U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

AD-A028 510

ADAPTIVE ESTIMATION AND PARAMETER IDENTIFICATION USING MULTIPLE MODEL ESTIMATION ALGORITHM

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

23 JUNE 1976

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ESD-TR-76-184	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Adaptive Estimation and Parameter Identification Using Multiple Model Estimation Algorithm		5. TYPE OF REPORT & PERIOD COVERED Technical Note 6. PERFORMING ORG. REPORT NUMBER Technical Note 1976-28
7. AUTHOR(s) Michael Athans and Chaw-Bing Chang		8. CONTRACT OR GRANT NUMBER(s) F19628-76-C-0002
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Project No. 8X363304D215
11. CONTROLLING OFFICE NAME AND ADDRESS Ballistic Missile Defense Program Office Department of the Army 1320 Wilson Boulevard Arlington, VA 22209		12. REPORT DATE 23 June 1976 13. NUMBER OF PAGES ~~X~~ 91
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Electronic Systems Division Hanscom AFB Bedford, MA 01731		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

multiple model estimation algorithm
multiple model filtering algorithm
multiple model prediction algorithm
multiple model smoothing algorithm
discrete time system
state estimation
Kalman filter algorithms

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this report is to introduce an adaptive estimation and parameter identification scheme which we shall call Multiple Model Estimation Algorithm (MMEA). Algorithms for filtering, smoothing, and prediction are derived for linear and nonlinear systems. Approaches for the extension of MMEA to a more general class of adaptive estimation problems are outlined. Several further research topics are also suggested.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

ADAPTIVE ESTIMATION AND PARAMETER IDENTIFICATION USING MULTIPLE MODEL ESTIMATION ALGORITHM

M. ATHANS
C. B. CHANG
Group 32

TECHNICAL NOTE 1976-28

23 JUNE 1976

LEXINGTON MASSACHUSETTS

Abstract

The purpose of this report is to introduce an adaptive estimation and parameter identification scheme which we shall call Multiple Model Estimation Algorithm (MMEA). The MMEA consists of a bank of Kalman filters with each matched to a possible parameter vector. The state estimates generated by these Kalman filters are then combined using a weighted sum with the a posteriori hypothesis probabilities as weighting factors. If one of the selected parameter vectors coincides with the true parameter vector, this algorithm gives the minimum variance state and parameter estimates. Algorithms for filtering, smoothing, and prediction are derived for linear and nonlinear systems. They are described in a tutorial fashion with results stated explicitly so that they can be readily used for computer implementation. Approaches for the extension of MMEA to a more general class of adaptive estimation problems are outlined. Several further research topics are also suggested.

Contents

Preceding page blank

I. INTRODUCTION

During the past decade considerable advances have been made in the theory, algorithms, and applications of stochastic estimation problems involving linear and nonlinear dynamics. The linear Kalman filter [1] and its diverse extensions to the nonlinear case [2,3,4] are well established theoretical and algorithmic tools with extensive applications.

In most practical applications of recursive estimation theory, there are difficulties in obtaining an exact mathematical model of the physical dynamic process. The uncertain parts of the system are sometime represented by an unknown parameter vector. Examples of this kind include the ballistic coefficient and lifting parameters modelled in the dynamics of a reentry vehicle [4,5,6,7,8]. When the state estimation for this type of system has to be carried out, the variations of these parameters and their identification play a critical role.

Many approaches have been proposed in attempting to perform state estimation together with parameter identification.* One very well-known on-line identification method is to model the unknown parameter as a Markov process with variance related to

*References in this category are too many to list, one may consult the IEEE Transactions on Automatic Control (Dec. 1974), a special issue on system identification, and reference [9] for listing of related references.

the system structure and the range of parameter variation. The restriction of this method is that its performance is critically influenced by the system structure, parameter variation, and the required bias and random errors. This technique usually works well within a rather small region of the state space and the variance of the process noise can only be determined by engineering intuition and extensive simulation study. This method however, has been able to produce excellent estimation accuracies in reentry vehicle tracking applications [5,6,8].

There exists an adaptive filtering and parameter identification method, which we shall call Multiple Model Estimation Algorithm (MMEA) in this report, which has attracted considerable attentions in the academic field [10, 11, 12, 13, 14]. This algorithm was first introduced by Magill [10] and later refined by Lainiotis [11]. The estimation algorithm was extended to adaptive control by Willner [12] and Upadhyay and Lainiotis [13].

The basic concept of MMEA is to construct a bank of Kalman filters with each matched to a possible parameter vector value. The state estimates generated by these Kalman filters are then combined using a weighted sum with the posteriori hypothesis probabilities as weighting factors. If one of the selected parameter vectors coincides with the true parameter vector, this method gives the minimum variance estimates of both the state vector and the parameter vector. In most physical problems, one usually has a

good idea of the possible values that a parameter may attain. Furthermore, the construction of the MMEA with a steady state Kalman filter bank requires only moderate computation. It therefore has attracted some attention for real-time applications [15, 16].

The purpose of this report is to introduce the Multiple Model Estimation Algorithm. It will be described in a tutorial fashion with results stated explicitly so that they can be readily used for computer implementation. Furthermore, the discussions on prediction and smoothing are believed to be new. Only the algorithms for discrete time system will be discussed. This is because that the modern estimation and control algorithms are mostly implemented on digital computers. Due to the fact that MMEA is theoretically more sound than the previous methods, it may be a potential candidate in trajectory re-construction applications.

This report is organized as follows. In the next section, the problem of state estimation with unknown parameters is formulated. Possible solutions are discussed in a tutorial fashion. In section three, the Multiple Model Filtering Algorithm (MMFA) is derived. The extensions to prediction (MMPA) and smoothing (MMSA) are presented in section four. Discussions of the first four sections assume linear system and measurement equations. The extension to the nonlinear system and methods of algorithm realization are presented in section five. A simple second order

example is included in section six to illustrate the theory. Discussions are given in the last section. Two appendices which list the linear smoothing algorithms and the Kalman and the extended Kalman filter algorithms are included for the reference purpose.

2. PROBLEM FORMULATION

2.1 Introduction

Consider a linear stochastic dynamic system whose dynamics depend on a parameter vector $\underline{\gamma}$. Let us write its equations in the standard state space representation and in the discrete time case.

State Dynamics

$$\underline{x}(t+1) = \underline{A}(\underline{\gamma})\underline{x}(t) + \underline{B}(\underline{\gamma})\underline{u}(t) + \underline{L}(\underline{\gamma})\underline{\xi}(t) \tag{2.1}$$

Measurement Equation

$$\underline{z}(t) = \underline{C}(\underline{\gamma})\underline{x}(t) + \underline{\theta}(t) \tag{2.2}$$

Next we define the different variables associated with eqs. (2.1) and (2.2).

The scalar t is a discrete time index

$$t = 0, 1, 2, \ldots. \tag{2.3}$$

The state vector $\underline{x}(t) \in R_n$ is an n-dimensional vector. The input or control vector $\underline{u}(t) \in R_m$ is an m-dimensional vector. The plant noise vector $\underline{\xi}(t) \in R_p$ is an p-dimensional vector. We assume that $\underline{\xi}(t)$ represents a zero mean discrete white noise sequence with known covariance matrix $\underline{\Xi}(t)$ - pxp matrix - i.e.

$$E\{\ \underline{\xi}(t)\ \} = \underline{0} \quad \text{for all } t \tag{2.4}$$

$$\operatorname{cov}[\ \underline{\xi}(t);\underline{\xi}(\tau)\ ] = E\{\ \underline{\xi}(t)\xi^T(\tau)\ \} = \underline{\Xi}(t)\delta(t,\tau) \tag{2.5}$$

where $\delta(t,\tau)$ is the Kroenecker delta

$$\delta(t,\tau) = \begin{cases} 1 & \text{iF} \quad t = \tau \\ 0 & \text{iF} \quad t \neq \tau \end{cases} \tag{2.6}$$

Note that the plant noise covariance matrix $\underline{\Xi}(t)$ is symmetric and at least positive semideninite

$$\underline{\Xi}(t) = \underline{\Xi}^T(t) \geqslant \underline{0} \tag{2.7}$$

The measurement noise vector $\underline{\theta}(t)\varepsilon R_r$ is an r-dimensional vector. We assume that $\underline{\theta}(t)$ represents a zero mean discrete white noise sequence with known covariance matrix $\theta(t)$ - an rxr matrix - i.e.

$$E \{ \theta(t) \} = \underline{0} \tag{2.8}$$

$$\text{cov} [\underline{\theta}(t);\underline{\theta}(\tau)] = E \{ \underline{\theta}(t)\underline{\theta}^T(\tau) \} = \underline{\theta}(t)\delta(t,\tau) \tag{2.9}$$

$$\underline{\Theta}(t) = \underline{\Theta}^T(t) > \underline{0} \tag{2.10}$$

Furthermore we assume that the plant driving noise $\underline{\xi}(t)$ and the measurement noise $\underline{\theta}(\tau)$ is independent for all values of t and τ, i.e.,

$$\text{cov} [\underline{\xi}(t);\underline{\theta}(\tau)] = \underline{0} \quad \text{for all } t,\tau \tag{2.11}$$

The above fix the dimensions of the different matrices that appear in eqs. (2.1) and (2.2). Thus

$\underline{A}(\gamma)$ is an nxn matrix

$\underline{B}(\gamma)$ is an nxm matrix

$\underline{L}(\gamma)$ is an nxp matrix

$\underline{C}(\theta)$ is an rxn matrix

2.2 The Parameter Vector γ

We have explicitly shown the dependence of the state dynamics and/or of the measurement equation upon the parameter vector $\underline{\gamma}$. We assume that the parameter vector $\underline{\gamma}\varepsilon R_q$ is a q-dimensional vector whose elements represent the key parameters.

The elements of the parameter vector $\underline{\gamma}$ are in general known only approximately. The degree of accuracy by which the elements of $\underline{\gamma}$ are known are strongly dependent upon the accuracy of modelling a physical process by eqs. (2.1) and (2.2) and the experiments that have been carried out.

In general, before the initiation of any real time estimation and/or control experiments, i.e.,prior to time t=0, one has some idea of the nominal value of the parameter vector, denoted by $\underline{\gamma}_o$, and of the degree of uncertainty (e.g.,standard deviations) associated with the nominal parameter values.

For the above reasons, it is reasonable to view the parameter vector $\underline{\gamma}$ as a random vector. All prior information about $\underline{\gamma}$ can be captured in its prior probability density function which we shall denote by $p(\gamma)$. At the very least, our best guess about $\underline{\gamma}$, prior to any additional real time experimentation,

is the nominal value $\underline{\gamma}_o$ which we can view as the unconditional prior mean

$$E\{\underline{\gamma}\} = \underline{\gamma}_o \tag{2.12}$$

The degree to which we "believe" the nominal value $\underline{\gamma}_o$ can be communicated to the mathematics by the prior covariance matrix $\underline{\Gamma}_o$ - a qxq matrix - of $\underline{\gamma}$, i.e.

$$\text{cov}[\underline{\gamma};\underline{\gamma}] = E\{(\underline{\gamma}-\underline{\gamma}_o)(\underline{\gamma}-\underline{\gamma}_o)^T\} \triangleq \underline{\Gamma}_o \tag{2.13}$$

It is also reasonable to assume that the uncertainty associated with the parameter vector $\underline{\gamma}$ has nothing to do with all other uncertainties. Thus we make the assumption

$$\begin{gathered}\underline{\gamma},\ \underline{x}(o),\ \underline{\xi}(t),\ \text{and}\ \underline{\theta}(\tau)\ \text{are independent}\\ \text{for all values of } t \text{ and } \tau\end{gathered} \tag{2.14}$$

2.3 The role of $\underline{\gamma}$ in Filtering Problems

First of all let us consider the filtering problem in the context of state estimation. To be more precise let us denote by the symbol Z(t) the total measurements obtained from the initial time T=0 to the present time t. These measurements include both the inputs applied to the system and the actual noisy sensor measurements. Thus if we assume that the first sensor measurement is carried out at t=1, and that the first input is applied at t=0, then the data set Z(t) is defined as follows

$$Z(t) = \{ \underline{z}(1), \underline{z}(2), \ldots, \underline{z}(t), \underline{u}(o), \underline{u}(1), \ldots, u(t-1)\} \quad (2.15)$$

In the state estimation version of the filtering problem one is interested in obtaining in real-time a "good" estimate of the actual value of the true state vector $\underline{x}(t)$ based upon the available data set $Z(t)$; this state estimate is commonly denoted by

$$\hat{\underline{x}}(t/t) \quad (2.16)$$

and the state estimation error is denoted by

$$\tilde{\underline{x}}(t/t) \triangleq \underline{x}(t) - \hat{\underline{x}}(t/t) \quad (2.17)$$

We can now have several cases, depending upon the relative uncertainty associated with the parameter vector $\underline{\gamma}$.

Case 1 Parameter vector known exactly

This is an unrealistic case and corresponds to the random vector $\underline{\gamma}$ having zero covariance

$$\underline{\Gamma}_o = \underline{0} \quad (2.18)$$

so that

$$\underline{\gamma} = \underline{\gamma}_o \quad (2.19)$$

Under these assumption, and the further assumption that all other random vectors, namely

$$\underline{x}(o), \underline{\xi}(t), \underline{\theta}(\tau)$$

are Gaussian, then the standard discrete time Kalman filter* [1] generates the optimal estimate of the state in the sense that the state estimate $\hat{\underline{x}}(t/t)$ is the true conditional mean of the state

$$\hat{\underline{x}}(t/t) = E\left\{ \underline{x}(t)/Z(t) \right\} \tag{2.20}$$

In addition one can calculate off-line, again through the discrete time Kalman filter algorithm the true conditional covariance matrix $\underline{\Sigma}(t/t)$

$$\underline{\Sigma}(t/t) = \mathrm{cov}\ [\ \underline{x}(t)\ ;\ \underline{x}(t)/Z(t)\ ] \tag{2.21}$$

Case 2 Parameter Uncertainty relatively "small"

In this case, we assume that the actual value of the parameter vector $\underline{\gamma}$ is "very close" to its nominal value. Thus, in this case, the parameter vector covariance matrix $\underline{\Gamma}_o$ is small.

$$||\ \underline{\Gamma}_o\ || = \text{small} \tag{2.22}$$

An alternate way of characterizing this is by

$$||\ \underline{\Gamma}_o\ || \ll ||\ \underline{\Xi}(t)\ ||,\ ||\ \underline{\Gamma}_o\ || \ll ||\ \underline{\Theta}(t)\ || \tag{2.23}$$

which means that the parameter uncertainty is much smaller than the uncertainty induced in the state by the plant noise $\underline{\xi}(t)$, and the errors introduced in the sensors by the measurement noise $\underline{\theta}(t)$. Under these circumstances, one can usually trust the robustness

*The discrete Kalman filter algorithm is stated in the Appendix A.

of the Kalman filter, as described in Case 1, to still generate "good" state estimates in the sense that

$$\underline{\hat{x}}(t/t) \simeq E \left\{ \underline{x}(t)/Z(t) \right\} \tag{2.24}$$

$$\Sigma(t/t) \approx \text{cov} \left[\underline{x}(t) ; \underline{x}(t)/Z(t) \right] \tag{2.25}$$

Case 3 Parameter Uncertainty Moderately low

As $||\underline{\Gamma}_o||$ increases, the errors of modelling the true values of the parameter vector $\underline{\gamma}$ by its nominal value $\underline{\gamma}_o$ become more significant and the performance of the standard Kalman filter begins to deteriorate. In this intermediate case, and especially when the major effect of the parameter uncertainty are reflected in the state dynamics (2.1), rather than the measurement equation (2.2), there have been several cures that have been suggested.

The basic rationale is that the increased parameter uncertainty in the system dynamics causes errors in the calculation of the one-step predicted estimate, $\underline{\hat{x}}(t + 1/t)$, of the standard Kalman filter algorithm. These errors can only be corrected by paying more attention to the measurements, which although noisy, still contain "good" information about the true state. Technically, this can be accomplished by increasing the magnitudes of the gains of the Kalman filter and, hence, the bandwidth of the Kalman filter.

One way of accomplishing this objective is to artificial-

ly increase selected elements of the plant noise covariance matrix $\underline{\Xi}(t)$. This trick has been often referred to as introducing fake white noise. If one can get away with it, in the sense that the state estimation errors $\underline{\tilde{x}}(t/t)$ remain acceptably small, then this procedure is desirable because one can still complete the (pseudo) covariance matrix Σ (t/t) and the Kalman filter gains off-line. However, this process of turning the Kalman filter is more of an art than a science.

The same philosophy of changing the magnitude of the plant noise covariance matrix $\underline{\Xi}(t)$, but on an on-line "adaptive" mode, is by monitoring the behavior of the residuals of the Kalman filter. The residual vector of time t, $\underline{r}(t/t)$, is defined as the difference between the actual measurement at time t, $\underline{z}(t)$, and the predicted measurement

$$\underline{r}(t/t) \triangleq \underline{z}(t) - \underline{C}(\underline{\gamma})\underline{\hat{x}}(t/t - 1) \tag{2.26}$$

In the case of no parameter uncertainty ($\underline{\Gamma}_o = \underline{0}$) the residuals are known to be zero-mean white and their covariance matrix, denoted by S(t/t), can be calculated from $\underline{\Sigma}(t/t)$. As the parameter uncertainty increases this is reflected in the nature of the residuals, in the sense that

(a) biases can be observed i.e.,

$$E\left\{ \underline{r}(t/t) \right\} \neq \underline{0} \tag{2.27}$$

(b) the residuals become correlated in time, so that they cease to be a white noise sequence.

A variety of methods that carry out real time tests in the residuals and subsequently change <u>on-line</u> the elements of the plant noise covariance matrix can be suggested. One of the simplest to implement is the one suggested by Jazwinski [2,17]. The price that one pays in these <u>adaptive filtering</u> methods is increased real-time computations associated with

(a) real-time tests and computations involving the residuals

(b) subsequent transformation of the residual-derived information into changes in the covariance matrix $\underline{\Xi}(t)$

(c) on-line calculations of the covariance equation and of the Kalman filter gain matrix

From a pragmatic point of view, these adaptive filtering algorithms change in a time-varying way the gains and the bandwidth of the Kalman filter, as modelling errors become significant and diagnosed in the residuals. If well designed, they can be effective in adjusting the bandwidth of the Kalman filter.

It should be noted that there is a tradeoff associated with high-gain, high-bandwidth Kalman filters. High-gain Kalman filters tend to decrease mean errors rapidly; on the other hand their high-bandwidth allows a greater amount of measurement noise

power to pass through, and this can cause increased RMS errors in the estimates. The successful prior timing and/or adaptive filtering algorithms have to take explicitly into account these mean errors vs. RMS errors tradeoffs.

Case 4 Moderate Parameter Uncertainty

As the parameter covariance matrix $\underline{\Gamma}_o$ increases further, the off-line or on-line turning of the basic Kalman filter cannot be counted upon to produce good estimation accuracy. This is due to the fact that the contributions of the parameter errors to model uncertainty can no longer be taken care of as equivalent white noise.

In such cases, one has to increase the real time complexity of the algorithm so as to explicitly carry out some on-line parameter estimation. In other words, in order to be able to obtain reliable state estimates, one has to obtain better estimates of the parameter vector $\underline{\gamma}$ based upon the real time measurements. In other words, the filtering algorithm has to simultaneously generate

(a) a state vector estimate, $\hat{\underline{x}}(t/t)$

(b) a parameter vector estimate, $\hat{\underline{\gamma}}(t/t)$.

Unfortunately, even in the simplest case, the joint state and parameter estimation problem constitutes a nonlinear filtering problem. It is well known, [18] to [22], that the true optimal solution to a nonlinear filtering problem, in the sense

of generating the true conditional mean of the state $E\{\underline{x}(t)/Z(t)\}$ requires the on-line solution of a set of non-linear partial differential equations at each and every time a measurement is made. For almost all problems of practical importance, the real time computational resources force the designer to use a suboptimal filtering algorithm.

The simplest suboptimal filtering algorithm is the so-called extended Kalman filter.* A slightly more complex algorithm is the so-called second order [4] or gaussian [2,23] filter.

The technique that is used to design the extended Kalman filter is that of state augmentation. Thus, in addition to eq. (2.1) which defines the dynamic stochastic evolution of the "natural" n state variables one writes another set of difference equations of the form

$$\underline{\gamma}(t+1) = \underline{\gamma}(t) \tag{2.28}$$

in case that it is known that the parameter vector $\underline{\gamma}$ is indeed a constant. If the parameter vector $\underline{\gamma}$ is known to change slowly with time, then the simplest way of modelling this is by the stochastic difference equation

$$\underline{\gamma}(t+1) = \underline{\gamma}(t) + \underline{\mu}(t) \tag{2.29}$$

*The extended Kalman filter algorithm is stated in the Appendix A.

where $\underline{\mu}(t)$ is a "fake" zero mean white noise process with covariance matrix

$$\mathrm{cov}\ [\ \underline{\mu}(t)\ ;\ \underline{\mu}(\tau)\ ] = \underline{M}(t)\delta(t,\tau) \tag{2.30}$$

The covariance matrix $\underline{M}(t)$ has to be suitably selected by the designer to reflect how rapidly and by how much one can reasonably expect the parameter $\underline{\gamma}$ to change or drift from its prior nominal value. We remark that more complex dynamic models than that shown in eq. (2.29) can be used if prior information on the "dynamics" of the parameter $\underline{\gamma}$ is available. The extended Kalman filter algorithm that generates the state estimate $\hat{\underline{x}}(t/t)$ and the parameter estimate $\hat{\underline{\gamma}}(t/t)$ has much more severe computational requirements than the algorithms discussed in Case 3. These additional requirements are due to the fact that at each measurement time one has to

(a) update an (n + q) - dimensional vector, the number (n) of state variables plus the number (q) of the parameters

(b) propagate an (n + q)x(n + q) (pseudo) covariance matrix using the standard extended Kalman filter covariance propagation formula.

(c) calculate a new (n + q)xr Kalman gain matrix

We remark that all the "tricks" discussed in Case 3 which involve the prior turning, or adaptive turning based on the residual be-

haviour, can be used in this case also to change the "fake white noise" covariance matrices $\underline{\Xi}(t)$ and $\underline{M}(t)$.

2.4 Discussion

The above brief semiphilosophical discussion points up some of the issues associated with the effects of uncertain parameters upon estimation problems. One can visualize the "robustness" of the varying complexity Kalman filters described in Cases 1 to 4 as shown in Figure 2.1

The way Figure 2.1 is to be interpreted is that if the true parameter is in band 3, then the estimators discussed in Cases 1,2 will not give satisfactory performance, while the estimators discussed in Case 3 will give good estimates. Needless to say the relative sizes or shapes of these robustness bands are next to impossible to calculate.

The point that we wish to stress, is that if the true parameter is outside the robustness band 4, then the extended Kalman filter discussed in Case 4 cannot be trusted to generate good state estimates, even though on-line parameter estimation is accomplished. The basic reason for this is that the covariance linearizations associated with the extended Kalman filter become invalid.

For this reason we shall explain in the next section how one can attack the problem of large parameter uncertainty through hypothesis testing and subsequently suggest a suboptimal procedure

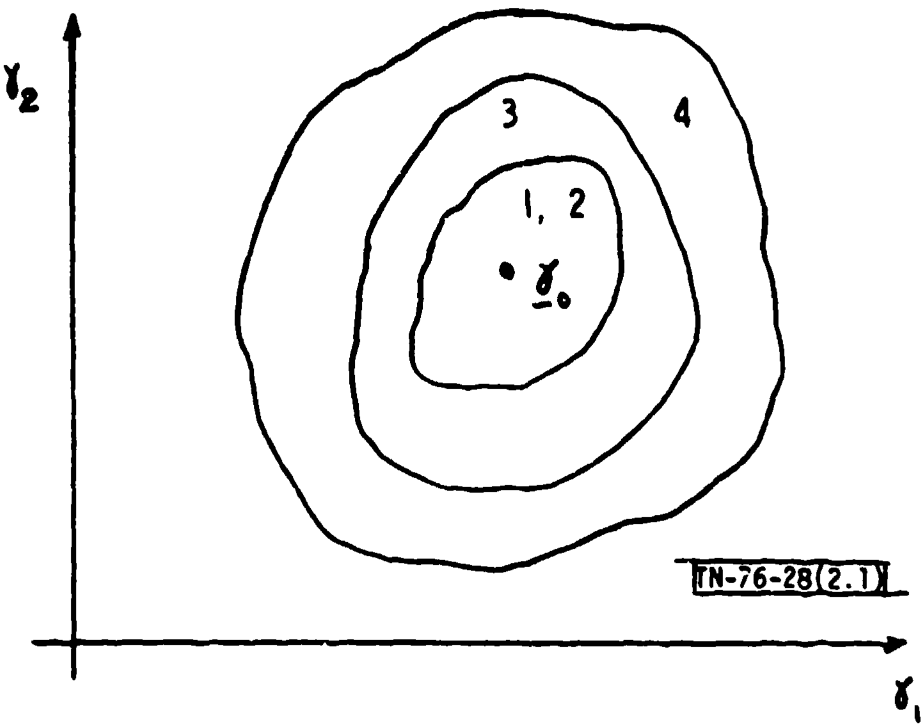

Fig. 2.1 Parameter uncertainty regions

that can be used for problems with large parameter uncertainty, as well as sudden transitions of the parameters (as it is the case with maneuvering reentry vehicles).

3. MULTIPLE MODELS FOR HYPOTHESIS TESTING AND STATE ESTIMATION: FILTERING

3.1 Introduction

In the previous section we have outlined the different methods that can be employed to carry out state estimation when the system dynamics contain uncertain parameters. We have concluded that as the parameter vector variance increases one is forced to employ nonlinear filtering algorithms, e.g., the extended Kalman filter, which simultaneously estimate the parameter vector and the desired state variables. We have also remarked that even these sophisticated algorithms will break down as the parameter uncertainty increases.

In this section we present the next most obvious level of complexity to take into account the effect of uncertain parameters. The first and simplest case is to subdivide the parameter space into regions and see what happens to the state estimation algorithm when such a discretization of the parameter space is carried out.

3.2 Discretization of the Parameter Space

As we have remarked in Section 2.2, the parameter vector $\underline{\gamma}$ is a q-dimensional vector. In most physical problems, one has some prior idea of the physical ranges of the elements of the parameter vector $\underline{\gamma}$. This engineering knowledge can be translated into a subset Ω_γ of R_q; the physical significance of Ω_γ is that it represents all reasonable values that the parameter vector $\underline{\gamma}$ can

attain.

The next step is to select a finite set of parameter values denoted by

$$\gamma_1, \gamma_2, \ldots, \gamma_N \tag{3.1}$$

These parameter vectors are scattered in the region Ω_γ.

3.3 Towards the MMFA; Assumptions

Let us suppose that the parameter vector $\underline{\gamma}$, which appears in the state space description of the stochastic dynamic system (2.1) - (2.2) does indeed coincide with one of the $\underline{\gamma}_i$ defined above. However, prior to making any measurements we do not know the "true index" i.

Needless to say, the above assumption is not true in any real life situation, in the sense that the true parameter vector $\underline{\gamma}$ will be "near," but not identical to, one of the $\underline{\gamma}_i$'s. Once more, we shall postpone discussion of this issue for the time being.

Under the assumption that indeed $\underline{\gamma}$ coincides with one of the $\underline{\gamma}_i$'s we can ask two questions:

1. What type of an algorithm can be used in order to generate

   a. the true conditional mean of the state, and

   b. the true conditional covariance matrix of the state

   given a set of past measurements. We remark that this constitutes the standard estimation or filtering question.

2. What type of an algorithm can be used to identify the true parameter $\underline{\gamma}_i$ given a set of past measurements. We remark that this constitutes an identification question.

One may argue that in many applications one may not be interested in the identification question, but only in the state estimation problem.* Nonetheless, it turns out that one cannot answer the questions independently, but one must obtain the answer to both questions simultaneously.

We shall next formulate the problem in a mathematically precise way, and then summarize the solution algorithm.

3.4 The MMFA: Formulation

For each value of $\underline{\gamma}$, denoted by $\underline{\gamma}_i$, let us redefine the matrices in section 2 as follows

$$A(\underline{\gamma}_i) \triangleq \underline{\underline{A}}_i, \quad \underline{B}(\underline{\gamma}_i) \triangleq \underline{\underline{B}}_i, \quad \underline{L}(\underline{\gamma}_i) \triangleq \underline{\underline{L}}_i \tag{3.2}$$

$$\underline{C}(\underline{\gamma}_i) \triangleq \underline{\underline{C}}_i \qquad ; \quad i = 1, 2, \ldots, N$$

We remark that the matrices $\underline{A}_i$, $\underline{B}_i$, $\underline{L}_i$, $\underline{C}_i$ can be time-varying; their time dependence is not explicitly shown.

*In the context of tracking RV's, if one tracks a ballistic RV, and the ballistic coefficient is viewed as the uncertain parameter, then one is usually interested in both state estimation for good tracking, and parameter estimation for discrimination. A similar situation exists for maneuvering re-entry vehicles; in the MaRV case one is interested in estimating parameters that are characteristic of the magnitude and direction of the maneuver accelerations.

Using the above notation, one has a class of N distinct linear stochastic dynamic systems described by

State Dynamics

$$\underline{x}(t+1) = \underline{A}_i\underline{x}(t)+\underline{B}_i\underline{u}(t)+\underline{L}_i\underline{\xi}(t) \quad ; \quad i=1,2,\ldots,N \qquad (3.3)$$

Measurement Equation

$$\underline{z}(t) = \underline{C}_i\underline{x}(t)+\underline{\theta}(t) \quad ; \quad i=1,2,\ldots,N \qquad (3.4)$$

The characteristics of the Gaussian plant noise $\underline{\xi}(t)$ are still given by eqs. (2.4) - (2.7), while the characteristics of the Gaussian measurement noise $\underline{\theta}(t)$ are still given by eqs. (2.8) - (2.11).

In addition to the plant noise, measurement noise, and initial state uncertainty, we must specify the parameter vector uncertainty. Under our assumptions, the random vector $\underline{\gamma}$ can attain a set of discrete values $\underline{\gamma}_1$, $\underline{\gamma}_2$, ..., $\underline{\gamma}_N$. In view of this, $\underline{\gamma}$ is a discrete random vector.

We can model this fact by a set of hypotheses. Let H be a scalar random variable (a hypothesis variable) and let

$$H_1, H_2, \ldots, H_N \qquad (3.5)$$

denote a set of events.

The interpretation that we attach to the event

$$H = H_j \quad \text{is} \quad \underline{\gamma} = \underline{\gamma}_j$$

and we can think of this phenomenon as that "nature" has selected the j-th linear system, defined by eqs. (3.3) and (3.4) and has placed it inside a black box.

Before we obtain any data from the system in the black box, we have to have some idea of the prior probability of which system is in the black box, or equivalently, the probability that $\underline{\gamma} = \underline{\gamma}_i$ for each i.

Let $\underline{P}_i(0)$ denote the prior probability that the i-th system is in the "black box." Thus

$$P_i(0) \triangleq \mathrm{Prob}(H=H_i) = \mathrm{Prob}(\underline{\gamma}=\underline{\gamma}_i) \tag{3.6}$$

with

$$P_i(0) \geq 0, \quad \sum_{i=1}^{N} P_i(0) = 1. \tag{3.7}$$

Thus, the probability density function, $p(H)$, of the random variable H is

$$p(H) = \sum_{i=1}^{N} P_i(0)\,\delta(H-H_i) \tag{3.8}$$

where $\delta(\cdot)$ is the Dirac delta function.

Remark: The numerical values of the prior probabilities $P_i(0)$ reflect to the mathematics our best guess on which models are more likely to be in the black box prior to their generating any data. If initially, i.e., at time t=0, any one of the models is equally likely, then we would select the $P_i(0)$ by

$$\underline{u}(0),\ \underline{u}(1),\ \ldots,\ \underline{u}(t-1) \tag{3.10}$$

and make a set of noise measurements

$$\underline{z}(1),\ \underline{z}(2),\ \ldots,\ \underline{z}(t) \tag{3.11}$$

from the system in the black box. As we have done in Section 2 we let $Z(t)$ denote the set of all past measurements

$$Z(t) = \{\ \underline{u}(0),\ \underline{u}(1),\ \ldots,\ \underline{u}(t-1),\ \underline{z}(1),\ \ldots,\ \underline{z}(t)\ \} \tag{3.12}$$

Define the probabilities

$$\begin{aligned} P_i(t) &\triangleq \mathrm{Prob}(H=H_i/Z(t)) \\ &= \mathrm{Prob}(\underline{\gamma}=\underline{\gamma}_i/Z(t)) \end{aligned} \tag{3.13}$$

to be the probability, given the measurement set $Z(t)$, that the i-th hypothesis (i.e., the i-th model) is the correct one. Clearly

$$P_i(t) \geq 0 \tag{3.14}$$

$$\sum_{i=1}^{N} P_i(t) = 1 \tag{3.15}$$

Given all of the above information and notation, we can list all the information that we would like to obtain, as well on the required algorithms to compute the variables of interest.

1. The <u>conditional</u> <u>mean</u> of the state

$$\hat{\underline{x}}(t/t) \underline{\Delta} E \{ \underline{x}(t)/Z(t) \} \tag{3.16}$$

2. The <u>conditional</u> <u>state</u> <u>covariance</u> matrix

$$\underline{\Sigma}(t/t) \underline{\Delta} \text{cov} [\underline{x}(t) ; \underline{x}(t)/Z(t)] \tag{3.17}$$

3. The dynamic evolution of the <u>posterior</u> <u>probabilities</u> $P_i(t)$; ideally we would like a recursive relation, i.e., $P_i(t+1)$ can be computed from the $P_j(t)$.

Remark: The conditional mean and the covariance can be computed once $p(x(t)/Z(t))$, the true conditional density function of the state of the system in the "black box" has been obtained.

3.5 The MMPA: Derivations

We shall obtain recursive relationships of the general conditional density functions at time t+1 given at time t.

We start by evaluating the conditional probability density function

$$p(\underline{x}(t+1)/Z(t+1)) \tag{3.18}$$

Use of the marginal density yields

$$p(\underline{x}(t+1)/Z(t+1)) = \int p(\underline{x}(t+1), H/Z(t+1))dH \tag{3.19}$$

From Bayes rule we obtain

$$p(\underline{x}(t+1),H/Z(t+1))=p(\underline{x}(t+1)/H,Z(t+1))p(H/Z(t+1)) \quad (3.20)$$

The conditional probability density $p(H/Z(t+1))$ can be written using the notation of eq. (3.13) as

$$p(H/Z(t+1)) = \sum_{i=1}^{N} P_i(t+1)\delta(H-H_i) \quad (3.21)$$

Substitute eqs. (3.20) and (3.21) into eq. (3.19), and integrate to obtain

$$p(\underline{x}(t+1)/Z(t+1)) = \sum_{i=1}^{N} P_i(t+1)p(\underline{x}(t+1)/H_i,Z(t+1)) \quad (3.22)$$

Remark: We know that the conditional densities $p(x(t+1)/H_i,Z(t+1))$ can be generated by a bank of N Kalman filters where each Kalman filter is "matched" to a distinct model, i.e., i-th hypothesis.

It is important to realize from basic Kalman filtering theory that the following relationship is true for each conditional probability density

$$p(\underline{x}(t+1)/H_i,Z(t+1)) = \frac{p(\underline{z}(t+1)/H_i,\underline{x}(t+1))p(\underline{x}(t+1)/H_i,Z(t))}{p(\underline{z}(t+1)/H_i,Z(t))} \quad (3.23)$$

and that

$$p(\underline{z}(t+1)/H_i,Z(t)) = \int p(\underline{x}(t+1)/H_i,\underline{x}(t))p(\underline{x}(t)/H_i,Z(t))\underline{dx}(t) \quad (3.24)$$

Remark: Under our assumptions all densities appearing in eqs. (3.23) and (3.24) are Gaussian, and hence they can be characterized by their mean and covariance matrix.

The basic idea is to construct a bank of N Kalman filters; each Kalman filter is designed using the specific parameter matrices $\underline{A}_i$, $\underline{B}_i$, $\underline{L}_i$, $\underline{C}_i$, $\underline{\Xi}$, $\underline{\Theta}$, and $\underline{\Sigma}_o$ (the initial state covariance matrix). Each Kalman filter in the bank is driven by the same input sequence, $\underline{u}(t)$, applied to the system in the "black box," and by the actual measurement sequence, $\underline{z}(t)$, generated by the system in the "black box."

Let $\hat{\underline{x}}_i(t/t)$ denote the state estimate generated by the i-th Kalman filter. More precisely, $\hat{\underline{x}}_i(t/t)$ is defined by.

$$\hat{\underline{x}}_i(t/t) \triangleq E\left\{ \underline{x}(t)/H_i,\ Z(t) \right\} = \int \underline{x}(t)p(\underline{x}(t)/H_i,\ Z(t))\underline{dx}(t) \quad (3.25)$$

Let $\underline{\Sigma}_i(t/t)$ denote the conditional covariance matrix associated with the i-th Kalman filter. More precisely

$$\underline{\Sigma}_i(t/t) \triangleq \text{cov}\left[\ \underline{x}(t);\underline{x}(t)/H_i,Z(t)\ \right]$$

$$= E\left\{ (\underline{x}(t)-\hat{\underline{x}}_i(t/t))(\underline{x}(t)-\hat{\underline{x}}_i(t/t))^T/H_i,Z(t) \right\}$$

$$= \int (\underline{x}(t)-\hat{\underline{x}}_i(t/t))(\underline{x}(t)-\hat{\underline{x}}_i(t/t))^T \cdot p(\underline{x}(t)/H_i,Z(t))\underline{dx}(t) \quad (3.26)$$

Remark: All the $\underline{\Sigma}_i(t/t)$, $i=1,2,\ldots N$ are precomputable.

In essence, from each Kalman filter mean $\hat{\underline{x}}_i(t/t)$ and covariance matrix $\underline{\Sigma}_i(t/t)$ we can construct the Gaussian density function $p(\underline{x}(t)/H_i,Z(t))$.

The next problem is to generate an overall estimate of the state, $\hat{\underline{x}}(t/t)$, according to eq. (3.16) of the system in the "black box." In addition, it is helpful to generate the true error covariance matrix, $\underline{\Sigma}(t/t)$, according to eq. (3.17), so that we have an idea of how accurate the estimate $\hat{\underline{x}}(t/t)$ of the true system state $\underline{x}(t)$ actually is.

We demonstrate below how the overall estimate $\underline{x}(t/t)$ can be generated once

a. The individual Kalman filter estimates $\hat{\underline{x}}_i(t/t)$ are available, and

b. The true conditional probabilities $P_i(t)$ defined by eq. (3.13) are available.

From eq. (3.22) we have

$$p(\underline{x}(t)/Z(t)) = \sum_{i=1}^{N} \underline{P}_i(t)p(\underline{x}(t)/H_i,Z(t)) \tag{3.27}$$

$$\begin{aligned}\hat{\underline{x}}(t/t) &= E\{\underline{x}(t)/Z(t)\} = \int \underline{x}(t)p(\underline{x}(t)/Z(t))\underline{dx}(t)\\ &= \sum_{i=1}^{N} P_i(t)\int \underline{x}(t)p(\underline{x}(t)/H_i,Z(t))\underline{dx}(t)\\ &= \sum_{i=1}^{N} P_i(t)\hat{\underline{x}}_i(t/t)\end{aligned} \tag{3.28}$$

Thus, the overall state estimate is the probabilistically weighted average, by the posterior (hypotheses) probabilities $P_i(t)$, of the state estimate generated by each one of the N Kalman filters.

To derive the true conditional covariance matrix $\underline{\Sigma}(t/t)$

we proceed as follows:

$$\underline{\Sigma}(t/t) \triangleq \text{cov}\ [\underline{x}(t);\underline{x}(t)/Z(t)]$$

$$= E\{(\underline{x}(t)-\hat{\underline{x}}(t/t)(\underline{x}(t)-\hat{\underline{x}}(t/t))^T/Z(t)\}$$

$$= \int(\underline{x}(t)-\hat{\underline{x}}(t/t))(\underline{x}(t)-\hat{\underline{x}}(t/t))^T p(\underline{x}(t)/Z(t))\underline{dx}(t)$$

$$= \sum_{i=1}^{N} P_i(t)\int(\underline{x}(t)-\hat{\underline{x}}(t/t))(\underline{x}(t)-\hat{\underline{x}}(t/t))^T \cdot$$

$$\cdot\ p(\underline{x}(t)/H_i,Z(t))\underline{dx}(t) \qquad (3.29)$$

After some algebra we obtain

$$\underline{\Sigma}(t/t) = \sum_{i=1}^{N} P_i(t)\ [\underline{\Sigma}_i(t/t)+(\hat{\underline{x}}_i(t/t)-\hat{\underline{x}}(t/t))\cdot$$

$$\cdot\ (\hat{\underline{x}}_i(t/t)-\hat{\underline{x}}(t/t))^T] \qquad (3.30)$$

Note that $\underline{\Sigma}(t)$ <u>cannot</u> be precomputed because it contains the real time estimates $\hat{\underline{x}}_i(t/t)$ generated by the Kalman filters in addition to the posterior probabilities $P_i(t)$ which as we shall see require real time measurements. The only remaining problem is to calculate dynamic evolution of the porbabilities $P_i(t)$

$$P_i(t) = \text{Prob}[H=H_i/Z(t)]$$

$$= \text{Prob}[\underline{\gamma}=\underline{\gamma}_i/Z(t)] \qquad (3.31)$$

We will relate each $P_i(t+1)$ to the $P_i(t)$ and other quantities that can be found from Kalman filters. The interesting

aspect of this calculation is that a truly recursive relationship can be obtained relating quantities only at successive measurement times, t and t+1, with relatively small computational burden.

Towards this goal we proceed as follows. Consider the conditional density

$$p(H/Z(t+1)) = \sum_{i=1}^{N} P_i(t+1)\,\delta(H-H_i) \tag{3.32}$$

Use of Bayes rule yields

$$\begin{aligned} p(H/Z(t+1)) &= p(H/\underline{z}(t+1),Z(t)) \\ &= \frac{p(H,\underline{z}(t+1)/Z(t))}{p(\underline{z}(t+1)/Z(t))} \\ &= \frac{p(\underline{z}(t+1)/H,Z(t))\,p(H/Z(t))}{p(\underline{z}(t+1)/Z(t))} \end{aligned} \tag{3.33}$$

But

$$p(H/Z(t)) = \sum_{i=1}^{N} P_i(t)\,\delta(H-H_i) \tag{3.34}$$

Note that according to our notation $Z(t+1) = \{Z(t),\underline{z}(t+1)\}$ Equations (3.32) to (3.34) yield

$$P_i(t+1) = \frac{p(\underline{z}(t+1)/H_i,Z(t))}{p(\underline{z}(t+1)/Z(t))}P_i(t) \tag{3.35}$$

The density $p(\underline{z}(t+1)/H_i,Z(t))$ is Gaussian and can be

calculated from the i-th Kalman filter

$$p(\underline{z}(t+1)/H_i, Z(t)) \sim N(\underline{C}_i(t+1)\hat{\underline{x}}_i(t+1/t), \underline{S}_i(t+1))^* \quad (3.36)$$

where

$$\underline{S}_i(t+1) = \underline{C}_i(t+1)\underline{\Sigma}_i(t+1/t)\underline{C}_i^T(t+1)+\underline{\theta}(t+1) \quad (3.37)$$

Note that the quantity $\underline{C}_i(t+1)\hat{\underline{x}}_i(t+1/t)$ is the predicted measurement at time t+1 generated by the i-th Kalman filter.

The matrix $\underline{S}(t+1)$ is the residual covariance matrix associated with the i-th Kalman filter. Note that the residual covariance matrices $\underline{S}_i(t+1)$ can be calculated off-line for each Kalman filter.

It remains to calculate the density $p(z(t+1)/Z(t))$ in eq. (3.35). Use of the marginal density leads to

$$\begin{aligned} p(\underline{z}(t+1)/Z(t)) &= \int p(\underline{z}(t+1), H/Z(t))dH \\ &= \int p(\underline{z}(t+1)/H, Z(t))p(H/Z(t))dH \\ &= \int p(\underline{z}(t+1)/H, Z(t)) \sum_{j=1}^{N} P_j(t)\delta(H-H_j)dH \\ &= \sum_{j=1}^{N} P_j(t)p(\underline{z}(t+1)/H_j, Z(t)) \end{aligned} \quad (3.38)$$

Remark: Once more all the densities $p(z(t+1)/H_j, Z(t))$ are available from the bank of Kalman filters; see eqs. (3.36)

*The notation $N(\underline{a}, \underline{A})$ denotes a Gaussian density with mean $\underline{a}$ and covariance $\underline{A}$.

and (3.37). Substituting eq. (3.38) into eq. (3.35) yields the desired result that the dynamic evolution of the probabilities $\underline{P}_i(t)$ is given by

$$P_i(t+1) = \frac{p(\underline{z}(t+1)/H_i,Z(t))}{\sum_{j=1}^{N} P_j(t)p(\underline{z}(t+1)/H_j,Z(t))} P_i(t) \tag{3.39}$$

where if we recall that

$$r = \dim \underline{z}(t) = \text{number of measurements} \tag{3.40}$$

then

$$p(\underline{z}(t+1)/H_j,Z\ (t)) = [2\pi]^{-\frac{r}{2}} [\det \underline{S}_j(t+1)]^{-\frac{1}{2}} \cdot$$

$$\cdot \exp\left\{-\frac{1}{2}(\underline{z}(t+1)-\underline{C}_j(t+1)\hat{\underline{x}}_j(t+1/t))^T S_j^{-1}(t+1)\right.$$

$$\left.\cdot\ (\underline{z}(t+1)-\underline{C}_j(t+1)\hat{\underline{x}}_j(t+1/t))\right\} \tag{3.41}$$

with

$$\underline{S}_j(t+1) = \underline{C}_j(t+1)\underline{\Sigma}_j(t+1/t)\underline{C}_j^T(t+1)+\underline{\theta}(t+1) \tag{3.42}$$

The relation (3.39) becomes more transparent if we introduce a somewhat simpler notation.

Let us define the residual (an r-dimensional vector) vector generated by each Kalman filter by

$$\underline{r}_i(t+1) \triangleq \underline{z}(t+1)-\underline{C}_i(t+1)\hat{\underline{x}}_i(t+1/t) \tag{3.43}$$

i.e., the difference between the actual measurement and the predicted measurement.

Then from each Kalman filter we can obtain the <u>scalar</u> quantity in real time

$$W_i(t+1) \triangleq \underline{r}_i(t+1)^T \underline{S}_i^{-1}(t+1)\underline{r}_i(t+1) \tag{3.44}$$

Also, let $\beta_i(t+1)$ denote the scalar <u>precomputable</u> quantity

$$\beta_i(t+1) \triangleq \left[2\pi\right]^{-\frac{r}{2}}\left[\det \underline{S}_i(t+1)\right]^{-\frac{1}{2}} \tag{3.45}$$

Using the above notation, the conditional density (3.41) can be written as

$$p(\underline{z}(t+1)/H_j, Z(t)) = \beta_i(t+1)\ \exp\left\{-\tfrac{1}{2}W_i(t+1)\right\} \tag{3.46}$$

From eqs. (3.46) and (3.39) we can now write the dynamic evolution of the probability density function as

$$P_i(t+1) = \frac{\beta_i(t+1)\exp\left\{-\frac{1}{2}\ W_i(t+1)\right\}}{\sum_{j=1}^{N} \beta_j(t+1)\exp\left\{-\frac{1}{2}\ W_j(t+1)\right\} P_j(t)}\ P_i(t) \tag{3.47}$$

The above formula illustrates that all measurements up to time t, Z(t), are captured in the posterior probabilities

$$P_1(t),\ P_2(t),\ \ldots,\ P_N(t) \tag{3.48}$$

The new measurement at time t+1, $\underline{z}(t+1)$, influence all

N residual vectors associated with the bank of Kalman filters according to eq. (3.43) and generate scalars $w_i(t+1)$, $i=1,2,\ldots,N$. This then can be used to update the probabilities

$$P_1(t+1),\ P_2(t+1),\ \ldots,\ P_N(t+1) \tag{3.49}$$

according to eq. (3.47). Thus, this represents a true recursive solution to the problem of probability updates.

A block diagram illustrating the MMFA is shown in figure 3.1.

3.6 The MMFA: Parameter Identification

In the previous subsection, we have described the basic idea of the Multiple Model Filtering Algorithm. In addition, we have derived algorithms for MMFA realization. In this subsection, we will show that the MMFA for parameter identification can be obtained in a straightforward manner. The minimum variance estimate of the unknown parameter $\underline{\gamma}$ is the conditional mean i.e.,

$$\hat{\underline{\gamma}}(t) = \int \underline{\gamma}\ p(\underline{\gamma}/Z(t))\,d\underline{\gamma} = E\{\underline{\gamma}/Z(t)\} \tag{3.50}$$

Recalling the fact that the events $H=H_i$ and $\underline{\gamma}=\underline{\gamma}_i$ are equivalent, we can rewrite eqn. (3.21) as

$$p(\underline{\gamma}/Z(t)) = \sum_{i=1}^{N} P_i(t)\,\delta(\underline{\gamma}-\underline{\gamma}_i) \tag{3.51}$$

where $\underline{P}_i(t)$ is interpreted as the probability that $\underline{\gamma}=\underline{\gamma}_i$ is true

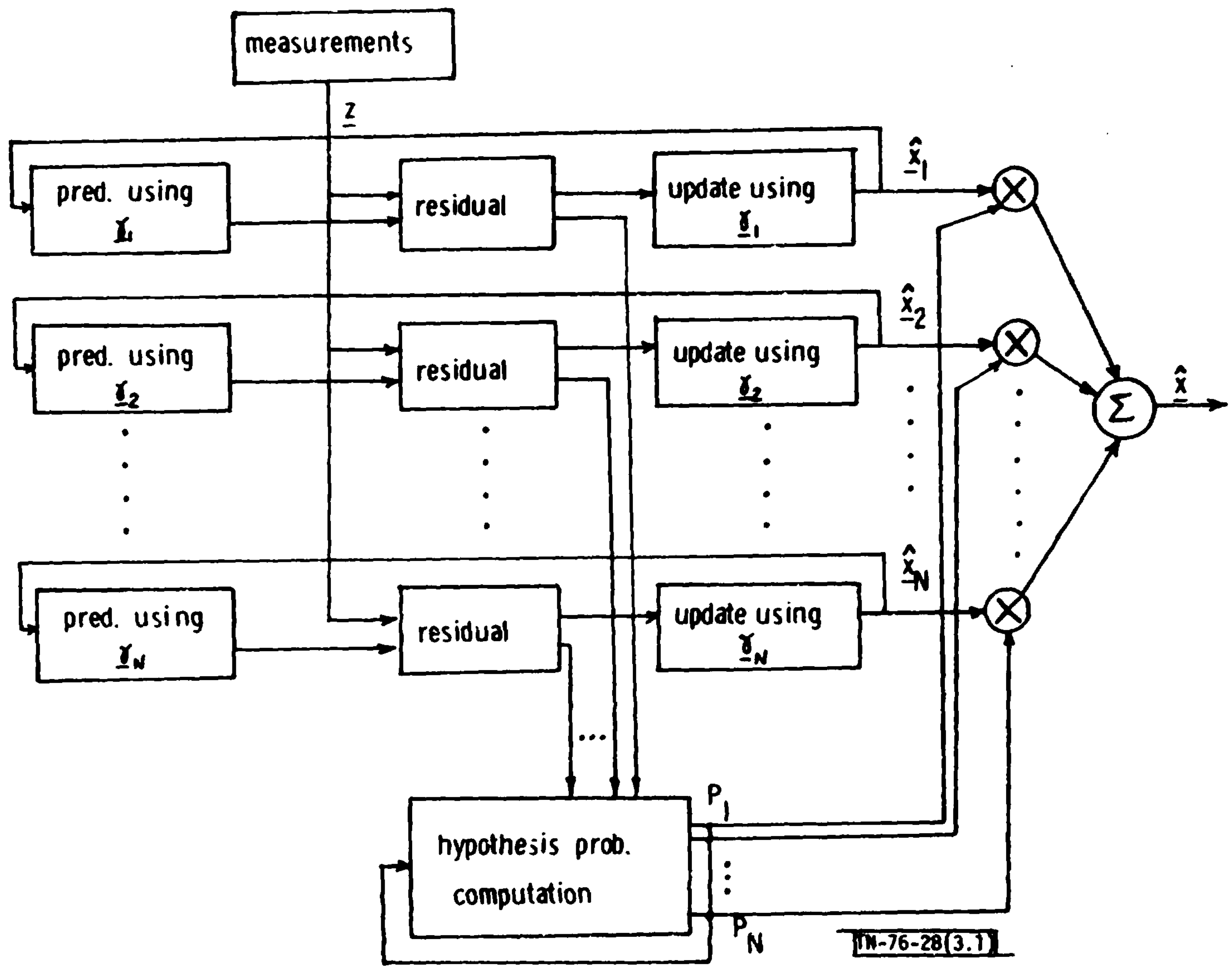

Fig. 3.1 The multiple model filtering algorithm block diagram

based upon all the data, $Z(t)$. Using (3.51) in (3.50) yields

$$\hat{\underline{\gamma}}(t) = \sum_{i=1}^{N} \underline{\gamma}_i \; P_i(t) \tag{3.52}$$

The covariance of $\hat{\underline{\gamma}}$ can be obtained similarly. Assuming that $\hat{\underline{\gamma}}$ is unbiased, then

$$\begin{aligned} \underline{\Sigma}_{\hat{\underline{\gamma}}}(t) &= \mathrm{cov}(\hat{\underline{\gamma}}(t)) \\ &= \int (\underline{\gamma} - \hat{\underline{\gamma}}(t))(\underline{\gamma} - \hat{\underline{\gamma}}(t))^T p(\underline{\gamma}/Z(t))\,d\underline{\gamma} \\ &= \sum_{i=1}^{N} P_i(t) \; [(\underline{\gamma}_i - \hat{\underline{\gamma}}(t))(\underline{\gamma}_i - \hat{\underline{\gamma}}(t))^T] \end{aligned} \tag{3.53}$$

3.7 Discussion

We now discuss the asymptotic properties of this algorithm from a heuristic point of view. If the system is subject to some sort of persistent excitation, then one would expect that the residuals of the Kalman filter associated with the correct model, say the i-th one will be "small", while the residuals of the mismatched filters ($j \neq i$, $j=1, 2, \ldots, N$) will be "large". Thus, if i indexes the correct model we would expect

$$W_i(t) \ll W_j(t) \qquad \text{for all } j \neq i \tag{3.54}$$

If such a condition persists over several measurements equation (3.47) shows that the "correct" probability $P_i(t)$ will increase while the "mismatched model" probabilities will decrease. To see this one can rewrite (3.47) as follows,

$$P_i(t+1) - P_i(t) = \left[\sum_{j=1}^{N} \beta_j(t+1)\exp\left\{-\frac{1}{2}W_j(t+1)\right\} P_j(t)\right]^{-1} \cdot$$

$$P_i(t)\left[(1-P_i(t))\beta_i(t+1)\exp\left\{-\frac{1}{2}W_i(t+1)\right\}\right.$$

$$\left. - \sum_{j\neq i} P_j(t)\beta_j(t+1)\exp\left\{-\frac{1}{2}W_j(t+1)\right\}\right] \qquad (3.55)$$

Under our assumptions

$$\exp\left\{-\frac{1}{2}W_i(t)\right\} \approx 1$$

$$\exp\left\{-\frac{1}{2}W_j(t)\right\} \sim 0.$$

Hence the correct probability will grow according to

$$P_i(t+1) - P_i(t) \approx \frac{P_i(t)[1-P_i(t)]\beta_i(t+1)}{\sum_{j=1}^{N} P_j(t)\beta_j(t+1)\exp\left\{-\frac{1}{2}W_j(t+1)\right\}} > 0 \quad (3.56)$$

which demonstrates that as $P_i(t) \to 1$, the rate of growth slows down.

On the other hand, for the incorrect model, indexed by $j\neq i$, the same assumptions yield

$$P_j(t+1) - P_j(t) \approx \frac{-P_j(t)P_i(t)\beta_i(t+1)}{\sum_{j=1}^{N} P_j(t)\beta_j(t+1)\exp\left\{-\frac{1}{2}W_j(t+1)\right\}} < 0 \qquad (3.57)$$

so that the probabilities decreased.

The same conclusions hold if we rewrite (3.47) in the form

$$P_i(t+1) - P_i(t) = \left[\sum_{j=1}^{N} P_j(t)\beta_j(t+1)\exp\left\{-\frac{1}{2}W_j(t+1)\right\}\right]^{-1}$$

$$\cdot P_i(t)\left[\sum_{j\neq i} P_j(t)\left(\beta_i(t+1)\exp\left\{-\frac{1}{2}W_i(t+1)\right\}\right.\right.$$

$$\left.\left.-\beta_j(t+1)\exp\left\{-\frac{1}{2}W_j(t+1)\right\}\right)\right] \tag{3.58}$$

The above discussion points out that this "identification" scheme is crucially dependent upon the regularity of the residual behavior between the "matched" and "mis-matched" Kalman filters.

As pointed out in reference [16], the dynamic evolution of the residuals may not follow the above regularity assumptions. This may be caused by errors in the selection of the noise statistics or using a steady state Kalman filter design, among others. To be specific, suppose that for a prolonged sequence of measurements the Kalman filter residuals turn out to be such that

$$W_1(t) \approx W_2(t) \approx \ldots \approx W_N(t) \tag{3.59}$$

Then

$$\exp\left\{-\frac{1}{2}W_i(t)\right\} \approx \alpha \quad \text{for all } i$$

Under this condition and using (3.58), we can see that

$$P_i(t+1) - P_i(t) \approx \frac{P_i(t) \sum_{j \neq i} P_j(t) \cdot (\beta_i(t+1) - \beta_j(t+1))}{\sum_{j=1}^{N} P_j(t) \beta_j(t+1)} \quad (3.60)$$

Suppose that one of the β_i's, say the β_k, is dominant, i.e.,

$$\beta_k > \beta_i \quad \text{for all } i \neq k$$

In this case, the right-hand side of (3.60) will be negative for all $i \neq k$, which means that all the $P_i(t)$ will decrease while the probability P_k (associated with the dominant β_k) will increase. The significance of this effect is that the β_i's are independent of the residuals and their magnitudes are not determined by which model is true. This issue, which has not been discussed in the literature, is believed to tie with the "identifiability" question of this scheme.

Above discussions merely point out possible shortcomings of this scheme. These issues may be adequately answered if we could address the following questions.

(1) A rigorous proof to show the asymptotic properties of the hypothesis probabilities. To the best of our knowledge, such a proof is not available in the literature.

(2) How would the hypothesis probabilities behave if none of the models coincide with the true model? Moor and Hawkes [14] used a distance measure to show that the probability associated with the model which is the closest one to the true model

will converge to unity. If this claim is warranted, one may be able to design an adaptive parameter discretization scheme which re-discretizes the parameter vector within the parameter subspace which is the closest to the true model as determined by the hypothesis probability and the distance measure.

(3) Answers to the above questions will certainly shade light to the identifiability problem.

Finally, let us re-emphasize the significance of this scheme from the estimation's point of view. This algorithm is <u>optimum</u> in the minimum variance sense in state and parameter estimation <u>if</u> the discretized parameter space indeed contains the true parameter. This is true because: (1) We use the conditional mean as the estimate and (2) the algorithm was derived without using any approximations.

4. MULTIPLE MODELS FOR HYPOTHESIS TESTING AND STATE ESTIMATION: SMOOTHING AND PREDICTION

4.1 Introduction

In the previous section, we have derived the multiple model filtering algorithm for state estimation when the system dynamics contain uncertain parameters. The parameter vector is discretized to cover a range of physical values that it may possibly attain. A bank of Kalman filters is built with each matching to a parameter vector. The a posteriori probability of a given model being true is used to combine the output of these filters. Algorithms for state estimation and parameter identification are derived.

In this section, the multiple modes smoothing and prediction algorithms (MMSA and MMPA) are derived.

4.2 The MMSA and MMPA: Assumptions

The system equations, measurement equations, parameter space, and hypothesis probability assumptions made in the section 3.4 are the same for the MMSA/MMPA derivation. We only modify the variables of interest to as follows:

1. The conditional mean of the state

$$\underline{\hat{x}}(\tau/t) \ \underline{\Delta} \ E \{ \underline{x}(\tau)/Z(t) \} \tag{4.1}$$

2. The conditional state covariance matrix

$$\underline{\Sigma}(\tau/t) \ \underline{\Delta} \ \text{cov} [\underline{x}(\tau);\underline{x}(\tau)/Z(t)] \tag{4.2}$$

3. The dynamic evolution of the posterior probabilities $P_i(t)$; again, we would like a recursive relation.

Remarks: (1) when $\tau>t$, it is called prediction.
when $\tau<t$, it is called smoothing.
when $\tau=t$, it is called filtering and this part of algorithm has already been presented.

(2) The conditional mean and the covariance can be computed once the conditional density function has been specified.

In the following, we re-state various forms of prediction and smoothing in terms of the evolution of $p(\underline{x}(\tau)/Z(t))$.

(1) Fix lag prediction/smoothing: update $p(\underline{x}(\tau)/Z(t))$ from $p(\underline{x}(\tau-1)/Z(t-1))$ where $\tau-t$ is a fixed constant

(2) Fix interval prediction/smoothing: update $p(\underline{x}(\tau)/Z(t))$ from $p(\underline{x}(\tau-1)/Z(t))$

(3) Fix point smoothing: update $p(\underline{x}(\tau)/Z(t))$ from $p(\underline{x}(\tau)/Z(t-1))$

4.3 The MMSA and MMPA: Derivations

Similarly, we start by evaluating the conditional probability density function

$$p(\underline{x}(\tau)/Z(t)) \tag{4.3}$$

Using the marginal density yields

$$\begin{aligned} p(\underline{x}(\tau)/Z(t)) &= \int p(\underline{x}(\tau),H/Z(t))dH \\ &= \int p(\underline{x}(\tau)/H,Z(t))p(H/Z(t))dH \end{aligned} \tag{4.4}$$

However

$$p(H/Z(t)) = \sum_{i=1}^{N} P_i(t)\delta(H-H_i) \tag{3.21}$$

and

$$P_i(t) = \text{Prob}(H=H_i/Z(t)) \tag{3.13}$$

Notice that $P_i(t)$ is interpreted as the probability of the event, $H=H_i$, being true conditioned upon all the measurements, $Z(t)$. Unlike the state and the covariance ((4.1) and (4.2)). <u>The hypothesis probability is only a function of one time variable, i.e., the time index of the measurement space</u>. Using (3.21) and (3.13) in (4.4) yields

$$p(\underline{x}(\tau)/Z(t)) = \sum_{i=1}^{N} P_i(t)p(\underline{x}(\tau)/H_i,Z(t)) \tag{4.5}$$

This equation is analogous to equation (3.22). Using (4.5), we obtain the predicted/smoothed state and covariance as

$$\begin{aligned}\hat{\underline{x}}(\tau/t) &= E\{\ \underline{x}(\tau)/Z(t)\ \} \\ &= \int \underline{x}(\tau)p(\underline{x}(\tau)/Z(t))d\underline{x}(\tau) \\ &= \sum_{i=1}^{N} P_i(t)\hat{\underline{x}}_i(\tau/t)\end{aligned} \tag{4.6}$$

$$\begin{aligned}\underline{\Sigma}(\tau/t) &= \text{cov}\ [\underline{x}(\tau)\ ;\ \underline{x}(\tau)/Z(t)] \\ &= \sum_{i=1}^{N} P_i(t)\int(\underline{x}/\tau) - \hat{\underline{x}}(\tau/t))(\underline{x}(\tau) - \hat{\underline{x}}(\tau/t))^T\end{aligned}$$

$$\cdot p(\underline{x}(\tau)/H_i, Z(t))d\underline{x}(\tau)$$

$$= \sum_{i=1}^{N} P_i(t)[\underline{\Sigma}_i(\tau/t) + (\hat{\underline{x}}_i(\tau/t) - \hat{\underline{x}}(\tau/t))$$

$$\cdot(\hat{\underline{x}}_i(\tau/t) - \hat{\underline{x}}(\tau/t))^T] \tag{4.7}$$

where $\hat{\underline{x}}_i(\tau/t)$ is the estimate from the i-th smoother/predictor and $\underline{\Sigma}_i(\tau/t)$ is the covariance of $\hat{\underline{x}}_i(\tau/t)$.

Remarks: (1) The realization of MMSA/MMPA again requires a bank of smoother/predictor with each matching to a possible parameter vector. The algorithms for the individual smoother/predictor realization have long been made available, for example, see [3, 24-28], or Appendix B.

(2) From the above derivation, the hypothesis probabilities $P_i(t)$ for smoothing/prediction are the same as those for filtering. The dynamic evaluation of $P_i(t)$ is still computed by using equation (3.47). Recalling that $P_i(t)$ is recursively updated by using the filter residuals. Since the filtering results at time t are obtained prior to any prediction and smoothing based upon $Z(t)$, the probabilities $P_i(t)$, $i=1, \ldots, N$ are always available.

(3) From equations (3.52) and (3.53), the parameter estimate is obtained as the weighted average of discretized parameter vectors. Again, there is only one time index which is the index of the measurement space. The smoothing/prediction algorithm for the parameter estimate is therefore the same as the filtering algorithm.

In summary, we state the following procedure for applying MMSA/MMPA.

(1) Compute filtering results, i.e., obtain $\hat{\underline{x}}_i(t/t)$, $\underline{\Sigma}_i(t/t)$, $P_i(t)$, $\hat{\underline{x}}(t/t)$, and $\underline{\Sigma}(t/t)$ from the algorithms of the previous section.

(2.a) For prediction, apply the individual predictor to

obtain $\underline{\hat{x}}_i(t+k/t)$ and $\underline{\Sigma}_i(t+k/t)$, i.e., iterate

$$\underline{\hat{x}}_i(t+1/t) = \underline{A}_i\underline{\hat{x}}(t/t) + B_i\underline{u}(t)$$

and

$$\underline{\Sigma}_i(t+1/t) = \underline{A}_i\underline{\Sigma}_i(t/t)\underline{A}_i^T + \underline{L}_i\underline{\Xi}(t)\underline{L}_i^T$$

k times with $\underline{\hat{x}}_i(t/t)$ and $\underline{\Sigma}_i(t/t)$ as initial conditions where k defines the discrete prediction time. The combined estimate $\underline{\hat{x}}(t+k/t)$ and covariance $\underline{\Sigma}(t+k/t)$ are obtained by using (4.6) and (4.7) with the hypothesis probabilities $P_i(t)$ the same as those obtained in step (1) (filtering).

(2.b) For smoothing, apply the individual smoother (see references [24-28] or Appendix B) to obtain $\underline{\hat{x}}_i(t-k/t)$ and $\underline{\Sigma}_i(t-k/t)$. The combined estimate $\underline{\hat{x}}(t-k/t)$ and covariance $\underline{\Sigma}(t-k/t)$ are obtained by using (4.6) and (4.7) while the hypothesis probabilities $P_i(t)$ are constant for all k and equal to those obtained in step (1).

5. MULTIPLE MODEL ESTIMATION ALGORITHM FOR NONLINEAR SYSTEMS

In this section, the MMEA for nonlinear systems is discussed. From the previous section, it is known that the smoothing and prediction are rather straightforward extensions of filtering, only the filtering algorithm will be emphasized here.

Similar to the linear case, we define the following nonlinear system and measurement equations corresponding to the i-th discretized parameter vector, $\underline{\gamma}_i$.

State Dynamics

$$\underline{x}(t+1) = \underline{f}(\underline{x}(t), \underline{u}(t), \underline{\xi}(t), \underline{\gamma}_i) \tag{5.1}$$

Measurement Equation

$$\underline{z}(t) = \underline{h}(\underline{x}(t), \underline{\theta}(t), \underline{\gamma}_i) \tag{5.2}$$

The plant noise $\underline{\xi}(t)$ is defined by equations (2.4) - (2.7) and the measurement noise $\underline{\theta}(t)$ is defined by equations (2.8) - (2.11).

The same as in the linear case, there are three separate steps in the multiple model estimation procedure, namely, the generation of individual state estimates matching to a given parameter vector, the evolution of the hypothesis probability and the combination of the individual estimates. Let each steps be discussed individually below.

(1) It is well-known that the realization of the optimum

state estimation for systems modelled by (5.1) and (5.2) involves solving a set of countably infinite differential equations [18 - 22]. It is therefore practically impossible to obtain these individual optimum estimates. Suboptimum filters will have to be used to construct the filter bank, i.e., to produce $\hat{\underline{x}}_i(t/t)$ approximately.

(2) The equation for updating the hypothesis probability is stated in equation (3.39)

$$P_i(t+1) = \frac{p(\underline{z}(t+1)/H_i, Z(t))}{\sum_{j=1}^{N} P_j(t)\, p(\underline{z}(t+1)/H_j, Z(t))} P_i(t) \qquad (3.39)$$

In arriving at this equation, no assumption was made on which type (linear or nonlinear) of systems are being considered. It is therefore still valid for nonlinear estimation. It however, cannot be calculated exactly due to the fact that the exact realization of the individual a posterior density $p(\underline{z}(t+1)/H_i, Z(t))$ can not be obtained. It can only be evaluated approximated with a sub-optimal filter (such as the extended Kalman filter*) for computing $\hat{\underline{x}}_i(t/t)$ and $\underline{\Sigma}_i(t/t)$.

(3) Assuming that the optimum individual estimate $\hat{X}_i(t/t)$ and its covariance $\underline{\Sigma}_i(t)$ are available, the optimum state estimate and its covariance can be computed by

*The extended Kalman filter equations are listed in the Appendix A.

$$\underline{\hat{x}}(t/t) = \sum_{i=1}^{N} P_i(t)\hat{\underline{x}}_i(t/t) \tag{3.28}$$

and

$$\underline{\Sigma}(t/t) = \sum_{i=1}^{N} P_i(t)\ [\underline{\Sigma}_i(t/t) + (\hat{\underline{x}}_i(t/t) - \hat{\underline{x}}(t/t)) \cdot (\hat{\underline{x}}_i(t/t) - \hat{\underline{x}}(t/t))^T] \tag{3.30}$$

Similarly, in order to realize (3.28) and (3.30) for states and measurements represented by (5.1) and (5.2), one has to use suboptimum filters to generate the individual estimates $\hat{\underline{x}}_i(t/t)$ and $\underline{\Sigma}_i(t/t)$.

Let us re-emphasize that equations (3.28), (3.29), and (3.30) are exact representations for the solution of the nonlinear estimation problem for systems modeled as (5.1) and (5.2). In other words, the a posterior hypothesis probabilities evolution and the method of computing the combined estimate are optimum if each individual estimate can be obtained optimally.

Numerous suboptimum filters have been proposed for nonlinear estimation [2,4,28-33]. The most popular filters are the extended Kalman filter and the second order filter [2,4] among others. Especially, the extended Kalman filter has attracted considerable attentions for practical applications [2-8]. The second order filter can generally provide improved performance

than the ext   K man filter with the trade-offs of higher computational burden. A comparison of various nonlinear filters may be found in [34,35]. All these filters may be used for the MMEA realization. A specific selection may be based on a particular physical problem and the required performance. For real-time application, one usually favors a simple filter pending on the available computer resources. For off-linear processing especially in the post-mission smoothing application, a sophisticated algorithm is usually preferred.

6. EXAMPLE

In this section, we present an example to illustrate the theory. Only the filtering algorithm is tested.

Consider the following second order continuous system.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -\gamma \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \tag{6.1}$$

This system may be used to describe the motion of a vehicle along a given axis with drag (represented by "γ") and control force (represented by "u"). If x_1 denotes the target range and a radar is used to take range measurements, the measurement equation is

$$z = x_1 + n \tag{6.2}$$

where n is measurement noise. The measurements are taken at discrete instance of times. A corresponding discrete system of (6.1) is

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & \frac{1-e^{-\gamma\Delta t}}{\gamma} \\ 0 & e^{-\gamma\Delta t} \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1-e^{-\gamma\Delta t}}{\gamma} \end{bmatrix} u \tag{6.3}$$

where Δt is the time interval between measurements. A multiple model filter is used to estimate x_1, x_2 and to identify γ. Three γ values are assumed, i.e., γ=0., .5, or 1.. The system and control matrices, $\underline{A}$ and $\underline{L}$ for those γ values with sampling interval

$\Delta t = .1$ are

(1) $\gamma = 0.$

$$\underline{A} = \begin{bmatrix} 1. & .1 \\ 0. & 1. \end{bmatrix} \quad ; \quad \underline{L} = \begin{bmatrix} 0. \\ .1 \end{bmatrix}$$

(2) $\gamma = .5$

$$\underline{A} = \begin{bmatrix} 1. & .0975 \\ 0. & .951 \end{bmatrix} \quad ; \quad \underline{L} = \begin{bmatrix} 0. \\ .0975 \end{bmatrix}$$

(3) $\gamma = 1.$

$$\underline{A} = \begin{bmatrix} 1. & .095 \\ 0. & .9 \end{bmatrix} \quad ; \quad \underline{L} = \begin{bmatrix} 0. \\ .095 \end{bmatrix}$$

The measurement noise standard deviation is equal to one. The time initial state is

$$x_1 = 100. \quad ; \quad x_2 = 50.$$

The following convention is used to relate the hypothesis to the parameter values.

$$H_1 \longrightarrow \gamma = 0.$$

$$H_2 \longrightarrow \gamma = .5$$

$$H_3 \longrightarrow \gamma = 1.$$

Two experiments are performed. They are described individually below.

Experiment 1: Parameter γ stays constant, control u is equal to zero.

Three cases with the true parameter being equal to one of the three possible values in each case are tested. The a posteriori hypothesis probabilities for all three cases are plotted in Figure 6.1. The initial hypothesis probabilities are uniformly distributed. The true system is always identified in within 10 data points

Experiment 2: Parameter γ jumps between models, control u is nonzero.

The control force is assumed to be equal to 50 and known to the estimator. Assuming the initial time is zero, the true γ time history is

$$\gamma = 0. \quad \text{for} \quad 0 \leq t < 2$$
$$\gamma = .5 \quad \text{for} \quad 2 \leq t < 4$$
$$\gamma = 1. \quad \text{for} \quad 4 \leq t < 6$$

It therefore represents a γ history with sudden jumps. The γ estimates are presented in Figure 6.2. Notice that the filter is always able to identify the true system. Two modifications are implemented in the algorithm in this case.

(1) The hypothesis probabilities are hard bounded. This is to prevent any probabilities from converging to zero (or

one). When it does, it will be very difficult for the probabilities to branch out again when the true system has actually switched. The bound used in this experiment is however, very small, i.e.,

$$Pr[H_i/Z(t)] \geq .0005 \quad \text{for} \quad i=1,2,3$$

(2) Although there is no process noise assumed in the system, a process noise term with covariance

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

is used in the filters. This is included also for the purpose of preventing the filter from being over-confident in its estimates therefore not able to switch to a different system. If there is no process noise added, the estimates of a mis-matched filter can drift far away from the true states. When the true parameter jumps to a different value, i.e., an originally mis-matched filter now becomes matched, it takes extremely long period of time for the algorithm to identify the true system again. Leaving proper process noise level in the filter will keep the mis-matched filter estimates sufficiently close to the true state so that the algorithm is adaptive to the parameter jumps. The control variable u also plays a critical role in this experiment. It represents a persistent excitation to explore differences among these systems. A basic issue which still needs answer is on the input design for system identification in using MMEA

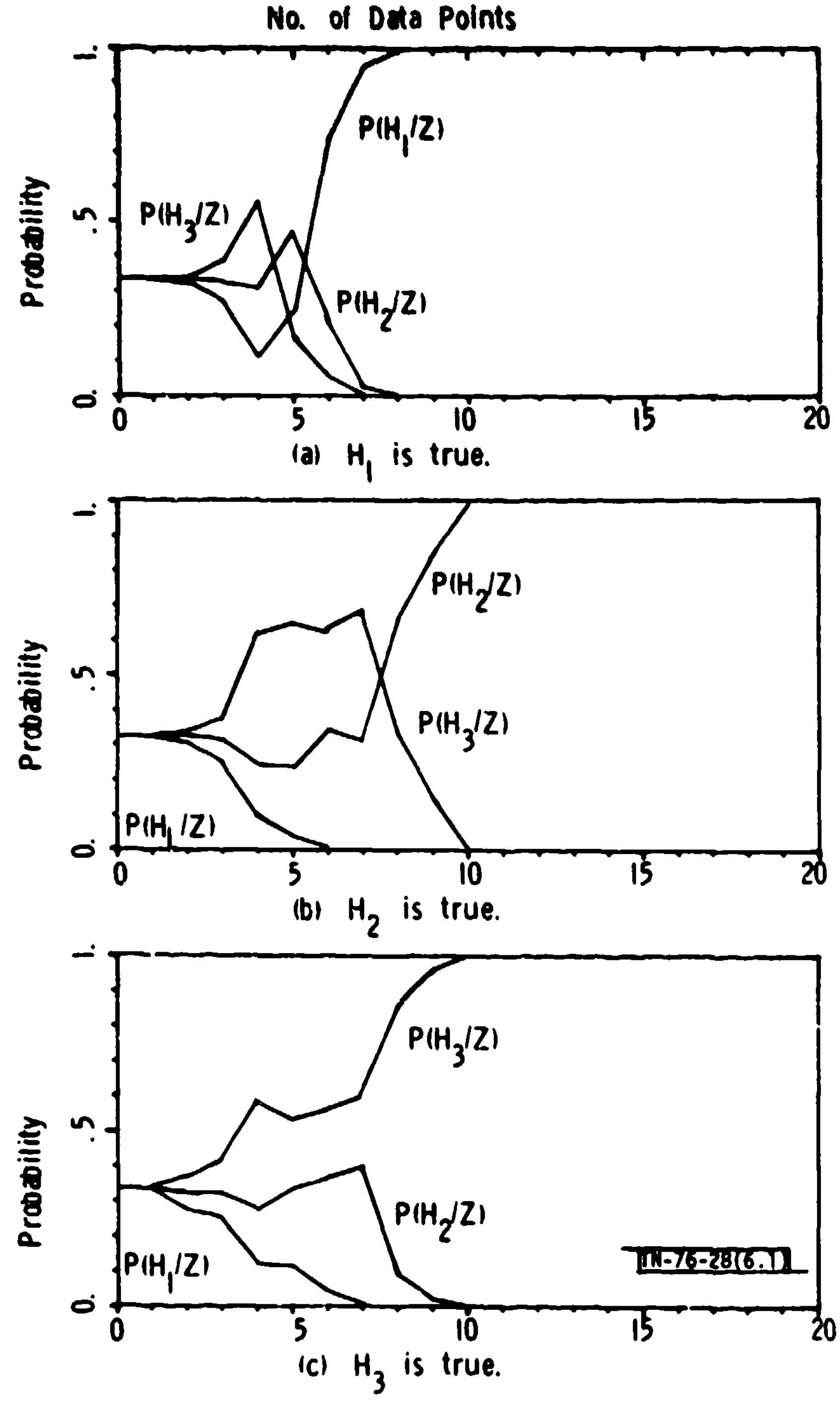

Fig. 6.1 Hypothesis probabilities of experiment 1

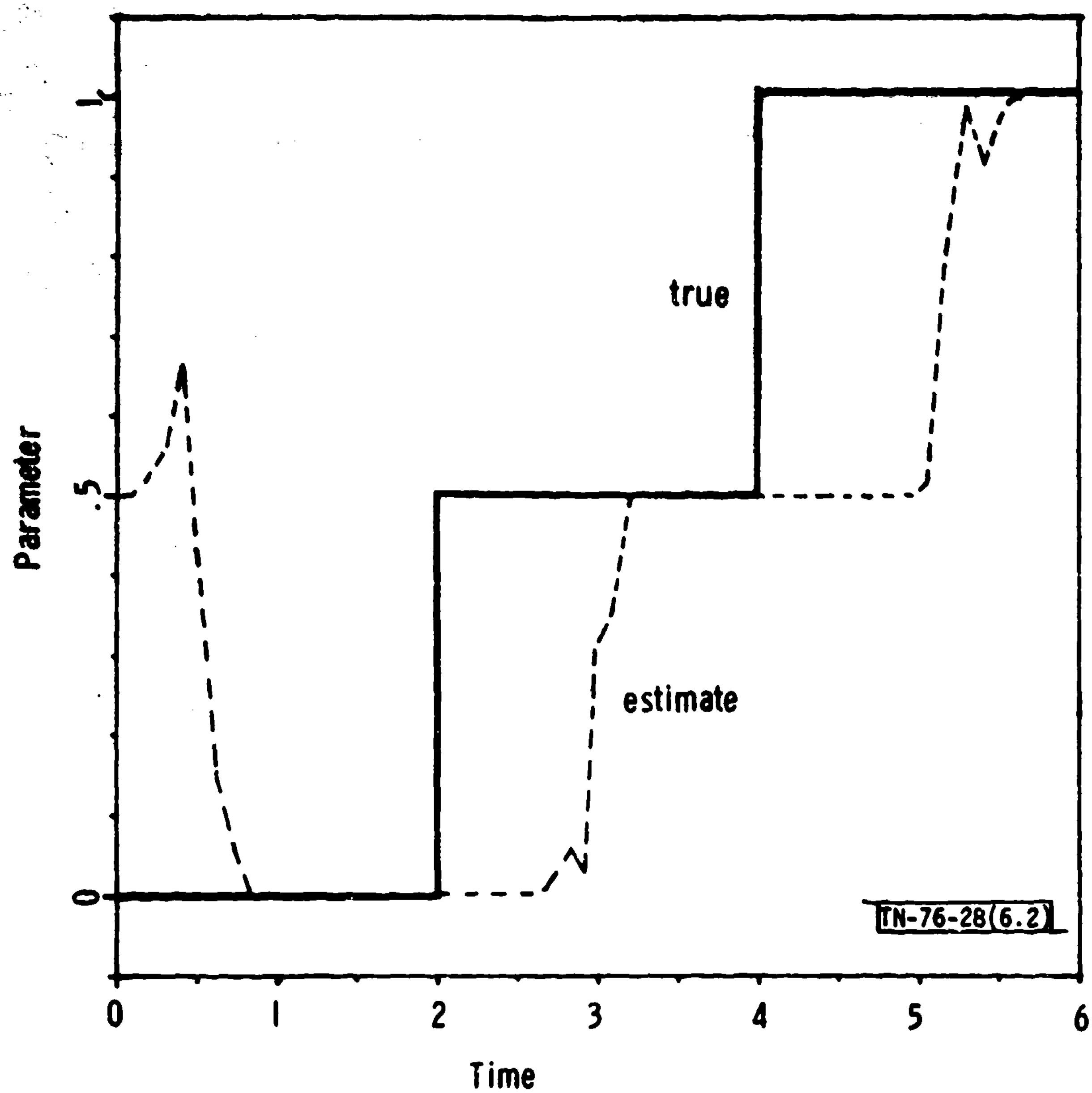

Fig. 6.2 Parameter estimates of experiment 2

method.

The above two experiments are simple but illustrative. The first experiment indicates that the MMFA can quickly identify the true system with a constant parameter. For time-varying parameters, some modifications are necessary so that the algorithm is adaptive to sudden parameter changes.

7. SUMMARY, DISCUSSION, AND FURTHER PROBLEMS

7.1 Summary

In this report, we have discussed the problem of state estimation with uncertain parameters and presented the solution by utilizating the multiple model estimation algorithm (MMEA). The following summaries pertaining to the properties of MMEA are listed without any specific order.

(1) Theoretically, the MMEA provides the minimum variance estimates of both state and parameter if one of the chosen models coincides with the true model.

(2) If the a posteriori hypothesis testing probabilities converge asymptotically, the true parameter is identified with probability one.

(3) The hypothesis probabilities for smoothing and prediction are the same as those for filtering.

(4) The hypothesis probability update equation and the weighted sum equations are optimum in the minimum variance sense and they are the same for both linear and nonlinear systems.

The usefulness of MMEA can only be fully understood and evaluated after applications to significant physical problems. Applications to the trajectory estimation area have still to be carried out. The application to the F-8C airplane real-time control system [16] has shown encouraging results and suggested further study areas in theory as well as in algorithm design.

7.2 Discussion: Extension to a Class of Time-Varying Parameters and Suboptimal Approaches

Strictly speaking, the MMEA presented in this report is optimum only for systems with time-invariant parameters. The theoretical and practical implications of using MMEA to systems with time-varying parameters are not completely understood. The example in the previous section has clearly indicated that some modifications must be incorporated in order to make the MMEA to follow parameter jumps. This is because once the true parameter is identified, the algorithm is locked on the true system and the mis-matched Kalman filter begins to drift away from the true state. When the true parameter has switched to a different value, it usually takes a long time for the algorithm to branch out again. The requirement for a time-varying parameter MMEA is to make the mis-matched output still sufficiently close to the true state and to keep the hypothesis probability from coming too close to zero (or unity).

There is a trivial extension of the MMEA to a special class of time-varying parameters. Consider the parameter space R_q which contains N parameter vectors each with dimension q, i.e.,

$$R_q = \{\underline{\gamma} : \underline{\gamma} = \underline{\gamma}_i \; ; \; i=1,\ldots,N\}$$

At the time t, the true parameter is equal to $\underline{\gamma}_i$. At the next instance of time, the true parameter may be equal to any parameters in R_q. As time progresses, the true parameter is changing

around with its values within R_q. Defining two types of hypotheses by

$H_i(t)$ = the hypothesis that $\gamma = \gamma_i$ at time = t is true, it is therefore a local hypothesis

$\bar{H}_k(t)$ = the hypothesis that a giving history for time up to t of γ indexed by k is true, it is therefore a global hypothesis.

These two types of hypothesis are related by the following equations

$$\bar{H}_k(t) = H_{k_t}(t) \oplus H_{k_{t-1}}(t-1) \oplus \text{----} \oplus H_{k_1}(1)$$

where the index for k_1, ..., k_t is 1, ..., N, the index for k is 1, ..., N^t, and $\oplus$ denotes the "and" operator. It is clear that each $\bar{H}_k(t)$ defines a possible sequence of γ history. With this definition, one may proceed in parallel to the development of this report to obtain a new MMEA for time-varying parameters. The derivation is briefly stated below.

1) For state estimate and covariance

Let

$$P_i(t) = \text{Prob}(\bar{H}(t) = \bar{H}_i(t)/Z(t)) \tag{7.1}$$

for i=1, ..., N^t. It is trivial to show that

$$\hat{\underline{x}}(t/t) = E(\underline{x}(t)/Z(t))$$

$$= \sum_{i=1}^{N^t} P_i(t)\hat{\underline{x}}_i(t/t) \tag{7.2}$$

$$\underline{\Sigma}(t/t) = \mathrm{cov}\,[\underline{x}(t),\ \underline{x}(t)/Z(t)]$$

$$= \sum_{i=1}^{N^t} P_i(t)\ [\underline{\Sigma}_i(t/t) + (\hat{\underline{x}}_i(t/t) - \hat{\underline{x}}(t/t)]$$

$$\cdot(\hat{\underline{x}}_i(t/t) - \hat{\underline{x}}(t/t))^T] \tag{7.3}$$

where $\hat{\underline{x}}_i(t/t) = E(\underline{x}(t)/\overline{H}_i(t), Z(t))$

$$\underline{\Sigma}_i(t/t) = \mathrm{cov}\ [\underline{x}(t), \underline{x}(t)\ /\ \overline{H}_i(t), Z(t)]$$

2) For probability update

Using the conditional probability relation yields

$$p(\overline{H}(t+1)/Z(t+1)) = p(H(t+1)/\overline{H}(t), Z(t+1))p(\overline{H}(t)/Z(t+1)) \tag{7.4}$$

Using Bay's rule on $p(H(t+1)/\overline{H}(t), Z(t+1))$ yields

$$p(H(t+1)/\overline{H}(t), Z(t+1))$$

$$= \frac{p(\underline{z}(t+1)/H(t+1), \overline{H}(t), Z(t))}{p(\underline{z}(t+1)/\overline{H}(t), Z(t))}\ p(H(t+1)/\overline{H}(t), Z(t)) \tag{7.5}$$

Define the following probability density functions

$$p(H(t+1)/\overline{H}(t),Z(t+1)) = \sum_{i=1}^{N} P(H_i(t+1)/\overline{H}(t) \quad ,Z(t+1))\,\delta(H-H_i) \quad (7.6)$$

$$p(H(t+1)/\overline{H}(t),Z(t)) = \sum_{i=1}^{N} P(H_i(t+1)/\overline{H}(t),Z(t))\,\delta(H-H_i) \quad (7.7)$$

where $P(H_i(t+1)/\overline{H}(t),Z(t+1)) = \text{Prob}(H(t+1)=H_i(t+1)/\overline{H}(t),Z(t+1))$ and $P(H_j(t+1)/\overline{H}(t),Z(t))$ is the probability that the parameter will switch to $\underline{\gamma}_i$ given a past history of $\underline{\gamma}$ and all the past measurements. It is determined by the property of the hypothesis process. If the hypothesis process is a Markov process, this probability becomes the transition probability, i.e.,

$$P(H_i(t+1)/\overline{H}(t),\ Z(t)) = P(H_i(t+1)/\overline{H}(t)) \quad (7.8)$$

$$= P(H_i(t+1)/H(t))$$

For example, if the parameter may change to any parameter in R_q with equal probability, we may assume

$$P(H_i(t+1)/\overline{H}(t),\ Z(t)) = \frac{1}{N} \qquad \text{for } i=1,\ \ldots,\ N.$$

Using (7.6) and (7.7) in (7.5) yields

$$P(H_i(t+1)/\overline{H}(t),Z(t+1))$$

$$= \frac{p(\underline{z}(t+1)/H_i(t+1),\overline{H}(t),Z(t))}{p(\underline{z}(t+1)/\overline{H}(t),Z(t))} P(H_i(t+1)/\overline{H}(t),Z(t)) \quad (7.9)$$

Using the equation

$$p(\overline{H}(t)/Z(t)) = \sum_{k=1}^{N^t} P_k(t)\delta(\overline{H} - \overline{H}_k) \quad (7.10)$$

in (7.9) yields

$$P(H_i(t+1)/\overline{H}_k(t),Z(t+1))$$

$$= \frac{p(\underline{z}(t+1)/H_i(t+1),\overline{H}_k(t),Z(t))}{p(\underline{z}(t+1)/\overline{H}_k(t),Z(t))} P(H_i(t+1)/\overline{H}_k(t),Z(t)) \quad (7.11)$$

where $p(\underline{z}(t+1)/H_i(t+1),\overline{H}_k(t),Z(t))$ is the residual density of the filter which was matched to the k-th history and is now matched to $\underline{\gamma}_i$ and

$$p(\underline{z}(t+1)/\overline{H}_k(t),Z(t))$$

$$= \sum_{m=1}^{N} p(z(t+1)/H_m(t+1),\overline{H}_k(t),Z(t))P(H_m(t+1)/\overline{H}_k(t),Z(t)) \quad (7.12)$$

Next, we relate $P(\overline{H}_i(t+1)/Z(t+1))$ to the conditional

probability. Using

$$P(H_i(t+1)/Z(t+1))$$

$$= \int P(H_i(t+1)/\overline{H}(t),Z(t+1))p(\overline{H}(t)/Z(t+1))d\overline{H}(t) \qquad (7.13)$$

and equation (7.10) one obtains the following equation for each $\overline{H}_k(t)$.

$$P(H_i(t+1),\overline{H}_k(t)/Z(t+1))$$

$$= P(H_i(t+1)/\overline{H}_k(t),Z(t+1))P(\overline{H}_k(t)/Z(t+1))$$

$$\text{for } i=1, \ldots, N \text{ and } k=1, \ldots, N^t \qquad (7.14)$$

where $P(H_i(t+1)/\overline{H}_k(t),Z(t+1))$ is specified in equation (7.11). Notice that $P(H_i(t+1),\overline{H}_k(t)/Z(t+1))$ is the updated hypothesis probability. Next, we derive the equation for computing $P(\overline{H}_k/Z(t+1))$. Using Baye's rule on $P(\overline{H}_k/Z(t+1))$ yields

$$P(\overline{H}_k(t)/Z(t+1))$$

$$= \frac{p(\underline{z}(t+1)/\overline{H}_k(t),Z(t))}{p(\underline{z}(t+1)/Z(t))} P(\overline{H}_k(t)/Z(t)) \qquad (7.15)$$

where $P(\overline{H}_k(t)/Z(t))$ is the a posteriori hypothesis probability at time = t, i.e., $P_k(t)$. The probability density functions of (7.15) are computed by using

$$p(\underline{z}(t+1)/\overline{H}_k(t),Z(t))$$

$$= \sum_{j=1}^{N} p(\underline{z}(t+1)/\overline{H}_k(t),H_j(t+1),Z(t))P(H_j(t+1)/\overline{H}_k(t),Z(t)) \tag{7.16}$$

and

$$p(\underline{z}(t+1)/Z(t))$$

$$= \sum_{m=1}^{N^t} p(\underline{z}(t+1)/\overline{H}_m(t),Z(t))P_m(t) \tag{7.17}$$

The probability update is therefore carried out by using equations (7.11), (7.14), and (7.15). These relations can be further condensed with the following simplified notations.

$$P(H_i/\overline{H}_k,Z(t)) = P(H_i(t+1)/\overline{H}_k(t),Z(t)) \tag{7.18}$$

$$\ell(H_i/\overline{H}_k) = \frac{p(z(t+1)/H_i(t+1),\overline{H}_k(t),Z(t))}{\sum_{j=1}^{N} p(\underline{z}(t+1)/H_j(t+1),\overline{H}_k(t),Z(t))P(H_j/\overline{H}_k,Z(t))} \tag{7.19}$$

$$= \text{conditional likelihood ratio}$$

$$\ell(\bar{H}_k) = \frac{p(\underline{z}(t+1)/\bar{H}_k(t), Z(t))}{p(\underline{z}(t+1)/Z(t))}$$

$$= \frac{\sum_{j=1}^{N} p(\underline{z}(t+1)/\bar{H}_k(t), H_j(t+1), Z(t)) P(H_j/\bar{H}_k, Z(t))}{\sum_{m=1}^{N^t} \{ \sum_{n=1}^{N} p(\underline{z}(t+1)/\bar{H}_m(t), H_n(t+1), Z(t)) P(H_n/\bar{H}_m, Z(t)) \} P_m(t)} \tag{7.20}$$

= likelihood ratio

Using (7.18), (7.19), and (7.20), equations (7.11), (7.14), and (7.15) may be combined to become

$$P_j(t+1) = \ell(H_i/\bar{H}_k) P(H_i/\bar{H}_k, Z(t)) \ell(\bar{H}_k) P_k(t) \tag{7.21}$$

Notice that for i=1, ..., N and k=1, ..., N^t, the index for j is 1, ..., N^{t+1}. The probability update is carried out with the conditional probability which characterizes the hypothesis process itself and the likelihood ratios which use the new information through residual density functions of each filter.

The MMEA for a constant parameter, i.e., the algorithm discussed in Section 3, is only a degenerate case of (7.21). When the parameter is a constant, the local hypotheses, H_i, and the global hypotheses, $\bar{H}_k$, become the same. The number of hypotheses is limited to the number of parameter vectors in R_q. Furthermore, the conditional probability of equation (7.11) becomes

$$P(H_i(t+1)/\bar{H}_k(t),Z(t)) \begin{cases} = 1 & \text{when} \quad H_i = \bar{H}_k = H_k \\ = 0 & \text{elsewhere} \end{cases} \tag{7.22}$$

Using (7.22) in $\ell(H_i/\bar{H}_k)$ and $\ell(\bar{H}_k)$ yields

$$\ell(H_i/\bar{H}_k) = \begin{cases} 1 & \text{when } H_i = \bar{H}_k = H_k \\ 0 & \text{elsewhere} \end{cases} \tag{7.23}$$

$$\ell(\bar{H}_k) = \frac{P(\underline{z}(t+1)/H_k(t+1),Z(t))}{\sum_{m=1}^{N} p(\underline{z}(t+1)/H_m(t+1),Z(t))P_m(t)} \tag{7.24}$$

Using (7.22), (7.23), and (7.24) in (7.21), we obtain equation (3.39), the probability update equation for the constant parameter case. This completes our a posteriori hypothesis probability derivation.

An obvious problem with this algorithm is that the number N^t, of $H_k(t)$, is growing with t. In order to make this algorithm practical, one has to limit the growing number of hypotheses. In the following, several suboptimal approaches for the time-varying parameter MMEA problem are outlined. The first two

approaches are aimed at limiting the number of possible sequences (or hypotheses). The last two approaches are mainly to reduce the chance of the algorithm being locked on a particular system.

(1) Maximum Likelihood Probability Approach

Consider the case that at time t there are only M hypotheses selected. For the next time period, each hypothesis may grow with N possibilities. It therefore has $M \cdot N$ hypotheses after each filter update. These $M \cdot N$ hypotheses are then limited by selecting only those M which have the largest hypothesis probabilities.

(2) Transition Probability and Finite Memory Hypothesis Process Approach

Suppose that the filtering process has limited memory so that $\bar{H}_k(t)$ is replaced by the most recent local hypothesis $H_k(t)$. Furthermore, it is assumed that the hypothesis process is a Markov process. Then one is interested in updating

$$P(H_i(t+1)/Z(t+1)) \qquad ; \quad i=1, \ldots, N.$$

from

$$P(H_i(t)/Z(t)) \qquad \text{for all } k=1, \ldots, N.$$

With this assumption and using (7.13), one obtains

$$P(H_i(t+1)/Z(t+1)) = \sum_{k=1}^{N} P(H_i(t+1)/H_k(t),Z(t+1))$$

$$\cdot\ P(H_k(t)/Z(t+1)) \qquad (7.25)$$

where $P(H_i(t+1)/H_k(t),Z(t+1))$ is obtained by an equation similar to (7.11), i.e.,

$$P(H_i(t+1)/H_k(t),Z(t+1))$$

$$= \frac{p(\underline{z}(t+1)/H_i(t+1),H_k(t),Z(t))}{p(\underline{z}(t+1)/H_k(t),Z(t))} P(H_i(t+1)/H_k(t),Z(t)) \qquad (7.26)$$

$P(H_k/Z(t+1))$ is obtained by an equation similar to (7.15), i.e.,

$$P(H_k/Z(t+1)) = \frac{p(\underline{z}(t+1)/H_k(t),Z(t))}{p(\underline{z}(t+1)/Z(t))} P(H_k(t)/Z(t)) \qquad (7.27)$$

where

$$p(\underline{z}(t+1)/H_k(t),Z(t))$$

$$= \sum_{j=1}^{N} p(\underline{z}(t+1)/H_k(t),H_j(t+1),Z(t))P(H_j(t+1)/H_k(t),Z(t)) \qquad (7.28)$$

$$p(\underline{z}(t+1)/Z(t)) = \sum_{m=1}^{N} p(\underline{z}(t+1)/H_m(t), Z(t))P_m(t) \quad (7.29)$$

Using (7.25), (7.26), and (7.27) one obtains an equation similar to (7.21), i.e.,

$$P_j(t+1) = P(H_j(t+1)/Z(t+1))$$

$$= \sum_{k=1}^{N} \ell(H_j/H_k)P(H_j/H_k)\ \ell(H_k)P_k(t) \quad (7.30)$$

where $P(H_j/H_k) = P(H_j(t+1)/H_k(t), Z(t))$

$=$ transition probability

$\ell(H_j/H_k) =$ conditional likelihood ratio defined in (7.26)

$\ell(H_k) =$ likelihood ratio defined in (7.27)

The difference of (7.21) and (7.30) is that with limited memory, we are interested in $P(H_j(t+1)/Z(t+1))$ and not in $P(\bar{H}_j(t+1)/Z(t+1))$. This also limits the number of filters to the number of $\underline{\gamma}_i$'s. One critical issue of this approach is the selection of the transition probability $P(H_j/H_k)$. In practical problems, it may be selected a priori with engineering intuition and physical reasons.

(3) Aging Filter Approach

When the system dynamics are uncertain and changing with time, the aging filter [36, 37] is often used to place exponentially higher weighting to the more recent measurements. Its extension to the MMEA case (e.g., in the probability computation) is not available. Preliminary results are discussed in [38].

(4) Others

There exist many methods that can be applied to open up the bandwidth of each Kalman filter and to prevent the a posteriori hypothesis probability from locking on zero (or unity). The method used in the previous section, i.e., increase process noise and bound the probability, is indeed just one of them.

A useful study would be to compare the above approaches by applying them to a significant physical problem, such as the Re-entry Vehicle Tracking problem.

7.3 Further Problem Areas

In this subsection, we conclude by suggesting the following further problem areas.

(1) From section 3.7, it was found that some fundamental issues of MMEA pertaining to its convergence and identifiability still require rigorous investigation.

(2) It is demonstrated in section 6 that a known input may be required in some situations to help identify time-varying parameters. The problem of optimal signal design in using MMEA for system identification is still an open issue.

(3) Further studies are required to extend MMEA to time-varying parameters. The optimum MMEA for a special class of time-varying parameters and several suboptimal approaches are discussed in section 7.2. The extension of MMEA to other types of parameter variation is needed. A Comparative study of the suboptimal approaches is an interesting further topic.

Acknowledgment

The authors would like to thank Dr. K. P. Dunn for many fruitful discussions.

APPENDIX A

THE DISCRETE KALMAN AND EXTENDED KALMAN FILTER ALGORITHMS

In this appendix, we state the discrete Kalman filter algorithm and its first order extension (the extended Kalman filter) to the nonlinear case.

A.1 The Discrete Kalman Filter Algorithm

Consider the discrete system represented by

$$\underline{x}(t+1) = \underline{A}\ \underline{x}(t) + \underline{B}\ \underline{u}(t) + \underline{L}\ \underline{\xi}(t) \tag{A.1}$$

with measurement equation represented by

$$\underline{z}(t+1) = \underline{C}\ \underline{x}(t+1) + \underline{\theta}(t+1) \tag{A.2}$$

where $\underline{x}$, $\underline{u}$, and $\underline{z}$ are state, control, and measurement vectors, respectively. $\underline{\xi}(t)$ and $\underline{\theta}(t)$ are white Gaussian noise sequences with zero mean and covariances $\underline{\Xi}(t)$ and $\underline{\Theta}$, respectively. The matrices, $\underline{A}$, $\underline{B}$, $\underline{L}$, and $\underline{C}$ may be time-varying although not explicitly shown. The discrete Kalman filter algorithm is stated below.

<u>Predict Cycle</u>

$$\hat{\underline{x}}(t+1/t) = \underline{A}\ \hat{\underline{x}}(t/t) + \underline{B}\ \underline{u}(t) \tag{A.3}$$

$$\underline{\Sigma}(t+1/t) = \underline{A}\ \underline{\Sigma}(t/t)\underline{A}^T + \underline{L}\ \underline{\Xi}(t)\underline{L}^T \tag{A.4}$$

Update Cycle

$$\hat{\underline{x}}(t+1/t+1) = \hat{\underline{x}}(t+1/t) + \underline{W}(t+1)(\underline{z}(t+1) - \underline{C}\,\hat{\underline{x}}(t+1/t)) \quad (A.5)$$

$$\underline{W}(t+1) = \underline{\Sigma}(t+1/t)\underline{C}^T[\underline{C}\,\underline{\Sigma}(t+1/t)\underline{C}^T + \underline{\theta}(t+1)]^{-1} \quad (A.6)$$

$$\underline{\Sigma}(t+1/t+1) = [\underline{I} - \underline{W}(t+1)\underline{C}]\,\underline{\Sigma}(t+1/t) \quad (A.7)$$

where

$$\hat{\underline{x}}(t/t) = E(\underline{x}(t)/Z(t)) \quad (A.8)$$

$$\hat{\underline{x}}(t+1/t) = E(\underline{x}(t+1)/Z(t)) \quad (A.9)$$

$$\underline{\Sigma}(t/t) = \mathrm{cov}(\underline{x}(t);\underline{x}(t)/Z(t)) \quad (A.10)$$

$$\underline{\Sigma}(t+1/t) = \mathrm{cov}(\underline{x}(t+1),\ \underline{x}(t+1)/Z(t)) \quad (A.11)$$

$Z(t)$ = the set of all past measurements

$$= \{\underline{u}(0),\ \underline{u}(1),\ldots,\underline{u}(t-1),\underline{z}(1),\ldots,z(t)\} \quad (A.12)$$

The initial estimate $\hat{\underline{x}}(0/0)$ is assumed to be Gaussian with mean $\underline{x}(0)$ and covariance $\underline{\Sigma}(0/0)$.

A.2 The Discrete Extended Kalman Filter Algorithm

Consider a nonlinear system represented by

$$\underline{x}(t+1) = \underline{f}(\underline{x}(t)) + \underline{B}\ u(t) + \underline{L}\ \underline{\xi}(t) \quad (A.13)$$

with measurement equation represented by

$$\underline{z}(t+1) = \underline{h}(\underline{x}(t+1)) + \underline{\theta}(t+1) \tag{A.14}$$

where all the matrices and vectors are the same as previously defined except that $\underline{f}(\ )$ and $\underline{h}(\ )$ now represent nonlinear system and measurement equations respectively. The extended Kalman filter is derived by expending $\underline{f}(\ )$ and $\underline{h}(\ )$ in using the Taylor series expansion up to first order term. Let

$$\underline{F} = \text{Jacobian matrix of } \underline{f}(\ )$$

$$= \left. \frac{\partial \underline{f}(\underline{x}(t))}{\partial \underline{x}(t)} \right|_{\underline{x}(t) = \hat{\underline{x}}(t/t)} \tag{A.15}$$

$$\underline{H} = \text{Jacobian matrix of } \underline{h}(\ )$$

$$= \left. \frac{\partial \underline{h}(\underline{x}(t+1))}{\partial \underline{x}(t+1)} \right|_{\underline{x}(t+1) = \hat{\underline{x}}(t+1/t)} \tag{A.16}$$

The discrete extended Kalman filter algorithm is stated below.

<u>Predict Cycle</u>

$$\hat{\underline{x}}(t+1/t) = \underline{f}(\hat{\underline{x}}(t/t)) + \underline{B}\, u(t) \tag{A.17}$$

$$\underline{\Sigma}(t+1/t) = \underline{F}\, \underline{\Sigma}(t/t)\underline{F}^T + \underline{L}\, \underline{\Xi}(t)\underline{L}^T \tag{A.18}$$

<u>Update Cycle</u>

$$\hat{\underline{x}}(t+1/t+1) = \hat{\underline{x}}(t+1/t) + \underline{W}(t+1)(\underline{z}(t+1) - \underline{h}(\hat{\underline{x}}(t+1/t))) \tag{A.19}$$

$$\underline{W}(t+1) = \underline{\Sigma}(t+1/t)\underline{H}^T[\underline{H}\, \underline{\Sigma}(t+1/t)\underline{H}^T + \underline{\Theta}(t+1)]^{-1} \tag{A.20}$$

$$\underline{\Sigma}(t+1/t+1) = [\underline{I} - \underline{W}(t+1)\underline{H}]\underline{\Sigma}(t+1/t) \qquad (A.21)$$

APPENDIX B

DISCRETE LINEAR SMOOTHING ALGORITHMS

The system and measurement equations are re-stated below.

$$\underline{x}(t+1) = \underline{A}(t+1,t)\underline{x}(t) + \underline{B}\ \underline{u}(t) + \underline{L}\ \underline{\xi}(t) \tag{B.1}$$

$$\underline{z}(t+1) = \underline{C}(t+1)\underline{x}(t+1) + \underline{\theta}(t+1) \tag{B.2}$$

All the definitions and statistical properties defined in the Appendix A still apply. Notice that the time-varying property of $\underline{A}(t+1,t)$ and $\underline{C}(t+1)$ is now explicitly shown. We still use the following definition for state estimate and covariance

$$\hat{\underline{x}}(\tau/t) = E[\underline{x}(\tau)/Z(t)] \tag{B.3}$$

$$\underline{\Sigma}(\tau/t) = \mathrm{cov}[\underline{x}(\tau);\underline{x}(\tau)/Z(t)] \tag{B.4}$$

Three kinds of smoothing are considered. They are defined below.

(1) Fixed-interval smoothing: given $Z(T)$, obtain $\hat{\underline{x}}(t/T)$ and $\underline{\Sigma}(t/T)$ for all $t<T$.

(2) Fixed-point smoothing: given τ, obtain $\hat{\underline{x}}(\tau/t)$ and $\underline{\Sigma}(\tau/t)$ for all $t>\tau$.

(3) Fixed-lag smoothing: advance $\hat{\underline{x}}(t/t+k)$ and $\underline{\Sigma}(t/t+k)$ to $\hat{\underline{x}}(t+1/t+1+k)$ and $\underline{\Sigma}(t+1/t+1+k)$ where k is a positive constant.

Only the algorithms will be stated here. Their derivations may be found in many references, e.g., refs [24-28]. These algorithms are stated individually in the following subsections.

B.1 Fixed-interval Smoothing Algorithm

In order to use the fixed-interval smoothing algorithm, the filtering results must be first made available.

State

$$\underline{\hat{x}}(t/T) = \underline{\hat{x}}(t/t) + \underline{G}(t)[\underline{\hat{x}}(t+1/T) - \underline{\hat{x}}(t+1/t)] \qquad \text{(B.5)}$$

Gain

$$\underline{G}(t) = \underline{\Sigma}(t/t)\underline{A}^T(t+1,t)\underline{\Sigma}^{-1}(t+1/t) \qquad \text{(B.6)}$$

Covariance

$$\underline{\Sigma}(t/T) = \underline{\Sigma}(t/t) + \underline{G}(t)[\underline{\Sigma}(t+1/T) - \underline{\Sigma}(t+1/t)]\underline{G}^T(t) \qquad \text{(B.7)}$$

Initial Conditions

$$\underline{\hat{x}}(T/T) \quad , \quad \underline{\Sigma}(T/T) \qquad \text{(B.8)}$$

B.2 Fixed-point Smoothing Algorithm

There are several equivalent algorithms in this category. Only one of them is stated here. Similary, the filtering results are needed for fixed-point smoothing.

State

$$\underline{\hat{x}}(\tau/t) = \underline{\hat{x}}(\tau/t-1) + \underline{D}(\tau/t)[\underline{\hat{x}}(t/t) - \underline{\hat{x}}(t/t-1)] \qquad \text{(B.9)}$$

Gain

$$\underline{D}(\tau/t) = \underline{D}(\tau/t-1)\underline{\Sigma}(t/t)\underline{A}^T(t+1,t)\underline{\Sigma}^{-1}(t+1/t) \quad \text{(B.10)}$$

Covariance

$$\underline{\Sigma}(\tau/t) = \underline{\Sigma}(\tau/t-1) - \underline{D}(\tau/t)\underline{W}(t)\underline{C}(t)\underline{\Sigma}(t/t-1)\underline{D}^T(\tau/t) \quad \text{(B.11)}$$

Initial Condition

$$\hat{\underline{x}}(\tau/\tau),\ \underline{\Sigma}(\tau/\tau),\ \underline{D}(\tau/\tau) = \underline{I} \quad \text{(B.12)}$$

where W(t) = filtering gain defined in (A.6).

$\underline{I}$ = identity matrix.

B.3 Fixed-lag Smoothing Algorithm

In order to perform fixed-lag smoothing, the filtering, fixed-interval smoothing, and fixed-point smoothing results must be available to obtain initial conditions.

State

$$\hat{\underline{x}}(t+1/t+1+k) = \underline{A}(t+1,t)\hat{\underline{x}}(t/t+k) + \underline{B}\,u(t)$$

$$+ \underline{L}\,\underline{\Xi}(t)\underline{L}^T\underline{A}^{-T}(t+1,t)\underline{\Sigma}^{-1}(t/t)[\hat{\underline{x}}(t/t+k) - \hat{\underline{x}}(t/t)]$$

$$+ \underline{D}(t+1/t+1+k)\underline{W}(t+1+k)[\underline{z}(t+1+k) - \underline{C}(t+1+k)\hat{\underline{x}}(t+1+k/t+k)] \quad \text{(B.13)}$$

Gain

$$\underline{D}(t+1/t+1+k) = \underline{G}^{-1}(t)\underline{D}(t/t+k)\underline{G}(t+k) \quad \text{(B.14)}$$

Covariance

$$\underline{\Sigma}(t+1/t+1+k) = \underline{\Sigma}(t+1/t) - \underline{D}(t+1/t+1+k)$$

$$\cdot \underline{W}(t+k+k)\underline{C}(t+1+k)\underline{\Sigma}(t+1+k/t)\underline{D}^T(t+1/t+1+k)$$

$$- \underline{G}^{-1}(t) \cdot [\underline{\Sigma}(t/t) - \underline{\Sigma}(t/t+k)]\underline{G}^{-T}(t) \qquad \text{(B.15)}$$

where W(t) = filtering gain, defined in (A.6). G(t) = fixed-interval smoothing gain, defined in (B.6).

Initial Conditions

$$\hat{\underline{x}}(t_o/t_o+k),\ \underline{\Sigma}(t_o/t_o+k),\ \underline{D}(t_o/t_o+k)$$

These conditions are obtained from fixed-point smoothing.

REFERENCES

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, J. Basic Eng., Ser. D, 82, 34-45, (1960).

[2] A. H. Jazwinski, Stochastic Processes and Filtering Theory (Academic Press, New York (1970).

[3] A. Gelb, ed., Applied Optimal Estimation (M.I.T. Press, Cambridge, Massachusetts, (1974).

[4] M. Athans, R. H. Whiting, and M. Gruber, "A Suboptimal Estimation Algorithm with Probabilistic Editing for False Measurements with Applications to Target Tracking with Wake Phenomena," to appear in IEEE Trans. Automatic Control.

[6] C. B. Chang, R. H. Whiting, and M. Athans, "On the State and Parameter Estimation for Maneuvering Re-entry Vehicles," to appear in IEEE Trans. Automatic Control.

[7] C. B. Chang, R. H. Whiting, and M. Athans, "Application of Adaptive Filtering Methods to Maneuvering Trajectory Estimation," Technical Note 1975-59, Lincoln Laboratory, M.I.T. (24 November 1975), DDC AD-B008137-L.

[8] M. Gruber "An Approach to Target Tracking," Technical Note 1967-8, Lincoln Laboratory, M.I.T. (10 February 1967), DDC AD-654272.

[9] M. Athans and P. P. Varaiya, "A Survey of Adaptive Stochastic Control Methods," Proc. Engineering Foundation Conference on Systems Engineering, New England College, Henniker, N.H., August 1975; also submitted to IEEE Trans. Automatic Control.

[10] D. T. Magill, "Optimal Adaptive Estimation of Sampled Stochastic Processes," IEEE Trans. Automatic Control AC-10, 434-439 (1965).

[11] D. G. Lainiotis, "Optimal Adaptive Estimation: Structure and Parameter Adaptation," IEEE Trans. Automatic Control AC-15, 160-170 (1971).

[12] D. Willner, "Observation and Control of Partially Unknown Systems," Ph.D. Dissertation, Department of Electrical Engineering, M.I.T.; also M.I.T. Electronic Systems Laboratory Report ESL-R-496 (June 1973).

[13] T. N. Upadhyay and D. G. Lainiotis, "Joint Adaptive Plant and Measurement Control of Linear Stochastic Systems," IEEE Trans. Automatic Control AC-19, 567-571 (1974).

[14] J. B. Moore and R. M. Hawkes, "Decision Methods in Dynamic System Identification," Proc. of 1975 IEEE Conf. on Decision and Control, Houston, Texas, pp. 645-650.

[15] M. Athans and D. Willner, "A Practical Scheme for Adaptive Aircraft Flight Control Systems," Proc. Symposium on Parameter Estimation Techniques and Applications in Aircraft Flight Testing, NASA TN D-7647, NASA Flight Research Center, Edwards, Calif. (April 1973), pp. 315-336.

[16] M. Athans, K. P. Dunn, C. S. Greend, W. H. Lee, N. R. Sandell, Jr., I. Segall, and A. S. willsky, "The Stochastic Control of the F-8C Aircraft Using the Multiple Model Adaptive Control (MMAC) Method," Proc. of 1975 IEEE Conf. on Decision and Control, Houston, Texas, pp. 217-228.

[17] A. H. Jazwinski, "Adaptive Filtering," Automatica 5, 475-485 (1969).

[18] H. J. Kushner, "On the Differential Equations Satisfied by Conditional Probability Densities of Markiv Processes with Applications," J. SIAM Control, Ser. A 2, 106-119 (1964).

[19] H. J. Kushner, "Dynamic Equations for Optimum Nonlinear Filtering," J. Differential Equations 3, 179-190 (1967).

[20] H. J. Kushner, "Nonlinear Filtering: the Exact Dynamical Equations Satisfied by the Conditional Mode," IEEE Trans. Automatic Control AC-12, 262-267 (1967).

[21] R. S. Bucy, "Nonlinear Filtering," IEEE Trans. Automatic Control AC-10, 198 (1965).

[22] M. Wonham, "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering," J. SIAM Control, Ser. A 2, 347-369 (1965).

[23] A. H. Jazwinski, "Nonlinear Filtering - Numerical Experiments," NASA - Goddard Astrodynamics Conf., Greenbelt, Maryland, April 1966.

[24] J. S. Meditch, Stochastic Optimal Linear Estimation and Control (McGraw-Hill, New York, 1969).

[25] J. S. Meditch, "Optimal Fixed Point Continuous Linear Smoothing," Proc. Joint Automatic Control Conf., June 1967, pp. 249-257.

[26] J. S. Meditch, "On Optimal Linear Smoothing Theory," Information and Control, 10, 598-615 (1967).

[27] J. S. Meditch, "A Successive Approximation Procedure for Nonlinear Data Smoothing," Proc. Symposium on Inform. Processing, Purdue University, Lafayette, Indiana, April 1969, pp. 555-568.

[28] A. P. Sage and J. L. Melsa, Estimation Theory with Applications to Communications and Control (McGraw-Hill, New York, 1971).

[29] B. Friedland and I. Bernstein, "Estimation of the State of a Nonlinear Process in the Presence of Nongaussian Noise and Disturbances," J. Franklin Institute 281, 455-480 (1966).

[30] H. Cox, "On the Estimation of the State Variables and Parameters for Noisy Dynamic Systems," IEEE Trans. Automatic Control AC-9, 5-12 (1964).

[31] V. O. Mowery, "Least Square Recursive Differential Correction Estimation in Nonlinear Problems," IEEE Trans. Automatic Control AC-10, 399-407 (1965).

[32] D. Detchmendy and R. Sridhar, "Sequential Estimation of States and Parameters in Noisy Nonlinear Dynamical Systems," Trans. ASME, J. Basic Engrg. 362-368 (1966).

[33] R. W. Bass, V. D. Norum, and L. Schwartz, "Optimal Multichannel Nonlinear Filtering," J. Math. Analysis and Applications 16, 152-164 (1966).

[34] L. Schwartz and E. B Stear, "A Computational Comparison of Several Nonlinear Filters," IEEE Trans. Automatic Control AC-13, 83-86 (1968).

[35] R. P. Wishner, J. A. Tabaczynski, and M. Athans, "A Comparison of Three Non-Linear Filters," Automatica 5, 487-496 (1969).

[36] T. J. Tarn and J. Zaborszky, "A Practical, Nondiverging Filter," AIAA J. 8, 1127-1133 (1970).

[37] R. W. Miller, "Asymptotic Behavior of the Kalman Filter with Exponential Aging," AIAA J. 9. 537-538 (1971).

[38] K. P. Dunn, "The Multiple Model Estimation Algorithm with Exponential Aging," to be issued,